

Rec'd PCT/PTO 10 DEC 2004

PCT/JP 03/07514

10/51754

日本国特許庁
JAPAN PATENT OFFICE

30.07.03

REC'D 19 SEP 2003

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application:

2002年 8月12日

出願番号
Application Number:

特願 2002-235294

[ST. 10/C]:

[JP 2002-235294]

出願人
Applicant(s):

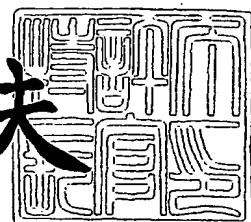
理化学研究所
株式会社ダナフォーム

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2003年 9月 5日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



BEST AVAILABLE COPY

出証番号 出証特 2003-3072722

【書類名】 特許願

【整理番号】 02771N

【特記事項】 特許法第36条の2第1項の規定による特許出願

【あて先】 特許庁長官殿

【国際特許分類】 C12N 15/00

【発明者】

【住所又は居所】 茨城県つくば市稲荷前22-1-201

【氏名】 林崎 良英

【発明者】

【住所又は居所】 埼玉県和光市広沢2-1 理化学研究所内

【氏名】 カルニンチ ピエロ

【発明者】

【住所又は居所】 東京都港区三田一丁目3-35 株式会社ダナフォーム
内

【氏名】 マティアス・ティ・ハーベス

【特許出願人】

【識別番号】 000006792

【氏名又は名称】 理化学研究所

【特許出願人】

【識別番号】 502049114

【氏名又は名称】 株式会社ダナフォーム

【代理人】

【識別番号】 100088546

【弁理士】

【氏名又は名称】 谷川 英次郎

【電話番号】 03(3238)9182

【先の出願に基づく優先権主張】

【出願番号】 特願2002-171851

【出願日】 平成14年 6月12日

【手数料の表示】

【予納台帳番号】 053235

【納付金額】 35,000円

【提出物件の目録】

【物件名】 外国語明細書 1

【物件名】 外国語図面 1

【物件名】 外国語要約書 1

【プルーフの要否】 要

【書類名】 外国語明細書

【発明の名称】 1 Title of Invention Method to utilize 5' ends of transcribed regions for cloning and analysis

【特許請求の範囲】

【請求項 1】 A method to prepare nucleic acids tags corresponding to the 5' end of transcribed regions said mRNA.

【請求項 1】 A method according to claim 1 where concatamers of such 5' end tags are produced.

【請求項 3】 A method in which such 5' end specific sequence tags derived from transcribed regions said mRNA are analyzed by sequencing.

【請求項 4】 The method for preparing concatamers of a plurality of at least two or more nucleic acid fragments having information on nucleotide sequences of 5' end regions of a plurality of nucleic acids related to transcribed in a sample, comprising

A first step of selectively collecting a plurality of cDNAs containing regions complementary to 5'-end regions of mRNAs, which cDNAs are formed by using RNA or mRNAs derived from a biological sample or in vitro synthesized RNA derived from cDNA - or tag - libraries in the sample as templates;

A second step to collecting fragments containing cDNA regions including at least the regions corresponding to the 5'-end regions of said mRNAs or cDNA;

And a third step of creating a concatamer of such 5' end nucleic tags.

【請求項 5】 The method as in claim 4 but in which

The first step is substituting the cap-structure of mRNAs with an oligonucleotide;

The second step constitute in the formation of full-length cDNA;

The third step involve cleavage of a 5' end tag and formation of concatamers.

【請求項6】 The method according to claim 4, wherein said first step comprises the steps of synthesizing the first-strand cDNAs using mRNAs as templates; attaching a selective binding substance to the cap structures of said mRNAs; cleaving single-stranded RNAs; binding said selective binding substance to a corresponding selective binding substance immobilized on a support, which corresponding selective binding substance selectively binds to said selective binding substance; and recovering said cDNA.

【請求項7】 A method as in claim 4 where the first step to isolate the full-length cDNA includes an RNase digestion step followed by treatment with an immobilized cap-binding substance followed by eluting such full-length cDNAs.

【請求項8】 A method to add a sequence connected to the 5' end a nucleic acids corresponding to the 5' terminal part of a transcript, when such that can be recognized by a substance that is capable of cleaving such nucleic acids outside the recognition sequence.

【請求項9】 The method according to claim 4, wherein said selective binding substance is biotin, and said corresponding selective binding substance is avidin, streptavidin or an avidin or streptavidin derivative which specifically binds to biotin.

【請求項10】 The method according to the claim 4 where the selective binding substance is digoxigenin and said corresponding binding substance is an antibody directed against digoxigenin.

【請求項11】 A method according to claim 4 or 9, wherein a selective binding substance is bound to a corresponding selective binding substance which is immobilized on to a support, and where such a support is made of magnetic beads, agarose beads, or latex beads

【請求項12】 The method according to any one of claims 4 and 6 to 11, wherein said second step comprises the steps of binding a linker ha

ving at least a restriction site for a substance that cleaves DNA outside its recognition sequence in the end region corresponding to the 5' end of said nucleic acids corresponding to the 5' end of genes, and a random oligomer region at the 3' end region; synthesizing a second-strand cDNA using said linker or other oligonucleotides partially or totally corresponding to the linker as a primer and said cDNA as a template; treating the obtained linker-bound double-stranded cDNA with said restriction enzyme; and selectively recovering fragments yielded by cleavage by the restriction enzyme, which fragments contain said linker moieties and part of 5' end cDNA.

【請求項13】 The method according to any of claims 4 to 12, where in a selective binding substance is attached to said linker; and the step of selectively recovering said fragments containing said linker moieties comprises the steps of binding said selective binding substance to a corresponding selective binding substance immobilized on a support, which corresponding selective binding substance selectively binds to said selective binding substance; and recovering said support.

【請求項14】 The method according to any of claims 4 to 13, where in said selective binding substance is biotin, and said corresponding selective binding substance is avidin, streptavidin, or an avidin derivative or derivatives of streptavidin which specifically binds to biotin.

【請求項15】 The method according to any of claims 4 to 13 where the selective binding substance is digoxigenin and said corresponding binding substance is an antibody directed against digoxigenin.

【請求項16】 The method according to any one of claims 4 to 15, wherein said restriction enzyme is a substance with an enzymatic activity to recognize nucleic acid and to cleave at a site different from the recognition site.

【請求項17】 The method according to any one of claims 4 to 16, w

herein said restriction enzyme is a class II restriction enzyme like *Gsu* I, *Mme*I, *Bpm* I or *Bsg* I

【請求項18】 The method using nucleic acid fragments obtained according to any one of claims 1 to 18, for further comprising the steps of cloning into concatamer.

【請求項19】 A method for determining nucleotide sequences of 5'-end regions of a plurality of mRNAs by sequencing said concatemer prepared by the method according to any one of claims 1 to 18.

【請求項20】 A method, which is the same method according to any one of claim 1 to 18, except that preliminarily obtained cDNAs having complete length is used instead of carrying out said first step.

【請求項21】 A method to produce 5' end nucleic acids tags corresponding to the 5' ends of mRNA, in which a mixture of RNA molecules is prepared from a preexisting full-length cDNA library and the obtained RNA carries at the 5' end of the RNAs a sequence cleavable by a substance able to recognize a nucleic acids and cleave outside its recognition sequence.

【請求項22】 A method to produce 5' end nucleic acids tags corresponding to the 5' ends of mRNA, in which a mixture of nucleic acids TAG molecules is prepared from a preexisting full-length cDNA library carrying close to the 5' end of a sequence cleavable by a substance able to recognize a nucleic acids and cleave outside its recognition sequence, which is used to produce a nucleic acid TAG molecule.

【請求項23】 The concatemer prepared by the method according to any one of claims 1 to 22.

【請求項24】 A vector comprising said concatemer according to claim 23.

【請求項25】 A sequence, which is derived from a concatemer prepared by the method according to any one of claims 1 to 22.

【請求項26】 A method based on any of claims 1 to 22, which allows to determine the transcriptional status of a given cell and therefore the transcriptional networking.

【請求項27】 A method, which is the same method according to any one of claims 1 to 22 to obtain expression data on a plurality of mRNA or cDNA in a sample.

【請求項28】 A method, which is the same method according to any one of claims 1 to 22, to quantify expression data on a plurality of mRNA in a sample.

【請求項29】 A method, which uses sequence information obtained from concatemers prepared by a method according to any one of claims 1 to 22, is used to build a database holding sequence information derived from the concatemers.

【請求項30】 A method, which is the same method according to any one of claims 1 to 22, to identify open reading frames in a genomic sequence said genome.

【請求項31】 A method, which is the same method according to any one of claims 1 to 22, to identify start sites of transcription and regulatory sequences upstream of the start site of transcription in a genomic sequence said genome.

【請求項32】 A method, which uses sequence information obtained from concatemers prepared by a method according to any one of claims 1 to 22, to clone a full-length or partial cDNA from a plurality of nucleic acids.

【請求項33】 A method, which uses sequence information obtained from concatemers prepared by a method according to any one of claims 1 to 22, to analyze the activity of regulatory regions in a genome said promoter.

【請求項34】 A method, which uses sequence information obtained f

rom concatemers prepared by a method according to any one of claims 1 to 22, to inactivate a gene.

【請求項35】 A method, which uses sequence information obtained from concatemers prepared by a method according to any one of the claims 1 to 22, to synthesis nucleotide sequences said linker.

【請求項36】 A method, which uses sequence information obtained from concatemers prepared by a method according to any one of the claims 1 to 22, to synthesize nucleotide sequences said primers.

【請求項37】 A method, which uses sequence information obtained from concatemers prepared by a method according to any one of the claims 1 to 22, to obtain extended nucleotide sequences derived from the 5'-ends of transcripts said sequencing.

【請求項38】 A method according to any one of claims 1 to 8, wherein a single stranded cDNA is ligated two a double stranded synthetic oligonucleotide said linker, wherein the linker has a single stranded overhang encompassing a nucleotide sequence said tag, which was obtained from concatemers prepared by a method according to any one of the claims 1 to 13, wherein the linker is attached to a selective binding substance and the selective binding substance is attached to a corresponding selective binding substance said support, and where such linker bound to the support is used to enrich a specific nucleotide sequence said 1st strand cDNA said RNA transcript.

【請求項39】 A method according to any one of claims 1 to 8, wherein a single stranded cDNA is ligated two a double stranded linker said primer, where a selective binding substance is attached to said linker, and where selectively binding substance is attached to a corresponding selective binding substance said support, and where such DNA template is used to obtain the nucleotide sequences of the 5' -region of an initial transcript said RNA.

【請求項40】 A method based on any of the claims 1-39 to be used for the development of diagnostic tools.

3 Detailed Description of Invention

【0001】

【発明の属する技術分野】

The present invention relates to a method to selectively collect multiple nucleic acid fragments containing information on the nucleotide sequences at the 5' end site of multiple mRNAs within a sample. The method of the present invention is effective for analyzing the mRNAs contained within the sample, for discovering new genes, and for studies on gene regulation.

【0002】

【従来の技術】

To utilize genomic information parts of the genome are transcribed into mRNA. For the understanding of the genome and its use in regulatory processes, information on individual mRNA species is required, which should include their partial or full-length nucleotide sequence and their relative or absolute quantity in a given biological context.

【0003】

Conventionally, the base sequences in mRNAs contained in a cell or tissue sample had been analyzed by preparing a cDNA library by reverse transcription, using mRNAs as templates and investigating the individual insert cDNA fragments within said cDNA library. Since a sample contains a large number of varied mRNAs, the conventional method is of limited efficiency to analyze gene expression profiles and to identify rare genes. Therefore other technologies have been invented to monitor the expression patterns of mRNA in complex samples and to identify genes by short sequence elements said tags.

【0004】

High-throughput expression profiling is commonly performed by the use of so-called DNA microarrays (Jordan B., DNA Microarrays: Gene Expression Applications, Springer-Verlag, Berlin Heidelberg New York, 2001; Schena A, DNA Microarrays, A Practical Approach, Oxford University Press, Oxford 1999). For such experiments specific probes representing individual genes or transcripts are placed on a support and simultaneously hybridized with a plurality of samples. Positive signals will be obtained where a probe on the support reacts with a molecule presented with the sample. These experiments allow the parallel analysis of a large number of genes or transcripts. However, the approach is limited to the fact that only genes or transcripts can be studied, which were initially identified by other experimental means. Such means can include cDNA libraries, partial sequence tags and/or results obtained from computer predictions. Due to the limitations of DNA microarray experiments alternative approaches are in use for gene discovery and expression profiling, which are based on partial sequences said tags obtained from a plurality of mRNA samples.

【0005】

The so-called SAGE (Serial Analysis of Gene Expression) method is known as an efficient method of obtaining partial information on the base sequences in mRNAs (Velculescu V.E. et al., Science 270, 484-487 (1995)). This method forms DNA concatamers by ligating multiple short DNA fragments (about 10 bp) containing information on the base sequences at the 3' end site of multiple mRNAs, and determines the base sequences in these DNA concatamers. It is a method for finding out partial information on the base sequences at the 3' end site of multiple mRNAs. When only a short base sequence close to the 3' end is available but the mRNA itself is already known, the SAGE method can often identify the mRNA, although the available base sequence is as short as about 10 bp. This method is currently in wide use as an important method for analyzing genes expressed

in specific cells or tissues.

【0006】

【発明が解決しようとする課題】

While the SAGE method can be used to learn a partial base sequence at the 3' end site of mRNAs, it is difficult to clone new genes based on the information in such short sequences at the 3' end site alone. Despite the application, SAGE does not teach how to obtain cDNA clones close to the 5' end of the cDNA. In fact, 4 bp restriction enzymes of class IIIs are used. A 4bp cutter usually cleaves on average a few hundred nucleotides, which is on average 1/10th of the average size of an mRNA transcript. Thus SAGE principles strongly suggest that 3' ends are collected with high prevalence, and no information can be collected about the 5' end for most of the transcript. In addition 10 bp tags have often been insufficient for specific gene identification and mapping to genomic sequences said entire or partial genomes. Therefore, the 10 bp tags are used to identify only a "sage-tag", which comprises a part of a mRNA. Notice that mammalian mRNA comprises only 3-5% of the transcribed part of the mammalian genome and the specific "sage-tag" comprises a subfraction of this 3-5%, which lies in proximity of the class IIIs restriction enzyme used in the analysis. Since a 4 bp restriction enzyme cuts approximately a random sequence every 4⁴ bp (256 bp), the "sage-tags" can represent approximately 1/256 of the 3-5% expressed fraction of the genome (calculation-less than 0.02%). Therefore, the SAGE techniques teach that essentially it is not possible to use SAGE-tags to analyze a genome but only a very limited fraction of it.

【0007】

Accordingly, the invention claimed in this application aims to provide new means that not only enables the acquisition of information on the base sequences of 5'-ends in mRNAs within a sample, but also enables the

cloning of new genes and the analysis of genomic sequence information, which correspond to coding and regulatory regions.

【0008】

This can include statistics on the DNA transcriptional starting site. By using concatamers to obtain information on a large number of 5'-sequence tags as presented in the invention, it is possible to effectively map transcriptional start sites and the related promoter sequences. Thus the invention provides new means, where SAGE did not allow any promoter analysis due to the use of unrelated 3'-ends. At the same time, there were techniques for the collection of full-length cDNA clones and sequences derived thereof; however, those are focusing on collecting the full-length cDNA clones and not fragments covering the 5'-ends. Therefore full-length cDNA cloning approaches are not suitable for high throughput identification and analysis of start sites of transcription and the related promoter regions. The invention offers here a novel way to combine contrasting teachings and to obtain by a high throughput approach 5' ends, which are useful for promoter mapping and analysis. The use of the invention to study and analyze complex regulatory networks in combination with the ability to identify and clone new genes opens a wide area of applications for the invention to monitor biological systems and their status in development, homeostasis, and disease.

【0009】

【課題を解決するための手段】

After devoted research, the inventors involved in this application were able to complete the present invention by arriving at the fact that by selectively collecting multiple nucleic acid fragments containing information on the base sequences at the 5' end site of the mRNAs, it is not only possible to acquire information on the base sequences in mRNAs, but it is also possible to clone new genes; and they were also able to arri

ve at a concrete method for attaining this goal.

【0010】

That is, the present invention provides a method for preparing concatemers of a plurality of nucleic acid fragments having information on nucleotide sequences of 5'-end regions of a plurality of mRNAs in a sample, comprising a first step of selectively collecting a plurality of first-strand cDNAs containing regions complementary to 5'-end regions of mRNAs, which cDNAs are formed by using mRNAs in the sample as templates; a second step of selectively collecting fragments containing cDNA regions including at least the regions complementary to the 5'-end regions of said mRNAs; and a third step of ligating the collected fragments to form a concatemer. The present invention also provides a method for determining nucleotide sequences of 5'-end regions of a plurality of mRNAs by sequencing said concatemer prepared by the method according to the present invention. The present invention further provides a method, which is the same method according to any one of claim 1 to 10, except that preliminarily obtained cDNAs having complete length is used instead of carrying out said first step. The present invention still further provides the concatemer prepared by the method according to the present invention. The present invention still further provides a vector comprising said concatemer according the present invention. The present invention still further provides sequence tags derived form said concatemers prepared according to the present invention. The present invention still further provides means to use the sequences derived from said concatemers to analyze the content of the plurality of a RNA sample. The present invention still further provides means to use the sequences derived from said concatemers to identify regions in the genome, which are required for gene regulation and gene expression.

【0011】

The invention is not limited to the use of concatamers for sequencing of 5' ends, but modifications at particular steps of the enrichment of 5' ends and their cloning as disclosed here allow for the individual sequencing of specific 5' ends. Such embodiments of the invention would include a modification of the first and second step, where a linker would be used that is specifically bound to a solid matrix. The cDNA bound to the support would then be used to prepare the sequencing reactions.

【0012】

Thus the invention refers more generally to the concept of isolating portions of nucleic acids corresponding to the 5' end of transcribed genes and using them to further high-throughput analysis such as sequencing

【0013】

【発明の実施の形態】

As described above, the method of the present invention can comprise but is not limited to roughly three steps each of which further comprises a plurality of steps. Each step will now be explained below. The concrete working examples of each step is described in detail in the later-mentioned working examples.

【0014】

Step 1

Step 1 is a step to selectively collect nucleic acids said cDNAs containing a site corresponding to the 5' end site of mRNAs within a sample and which are synthesized for instance by using said mRNAs as templates.

【0015】

Either total RNA or mRNA taken from a desired cell or tissue can be used as the starting substrate. The preparation method of total RNA and mRNA is already known, and it is also described in detail in the later-mentioned working examples. In other embodiments, a full-length cDNA librar

y may be used to isolate the 5' end nucleic acids corresponding to the 5' end of the transcribed part of the genes. Alternatively, a cDNA library itself would be cleaved if it carries a Class II S enzyme in proximity of the 5' end.

【0016】

Step 1 itself can be conducted by a publicly known method. In other words, methods to construct full-length cDNAs and methods to synthesize cDNA fragments at least containing a site corresponding to the 5' end site of the mRNAs are already known, and any of these methods can be adopted. One of the preferable methods is the cap trapper method (e.g. Piero CARNINCI et al., METHODS IN ENZYMOLOGY, VOL. 303, pp. 19-44, 1999). This cap trapper method shall be explained below, however, the invention is not limited to the use of the cap trapper method and other approaches to enrich or select full-length cDNAs could be applied as well. An alternative method (as described by Pelletier et al. in 1995) makes use of an immobilized cap-binding protein to isolate full-length cDNAs after RNase treatment of a hybrid.

【0017】

Alternatively to the cap-selection, one could dephosphorylate with a phosphatase, such as BAP (bacterial alkaline phosphatase) the 5' ends of mRNAs, followed by treatment with the decapping enzyme TAP (tobacco acid pyrophosphatase). Subsequently a ribonucleotide or a deoxyribonucleotide can be attached to the 5' end of the mRNA instead of the original cap-structure with RNA ligase (Maruyama and Sugano). In this way, for instance, a Class II S recognition site could be placed on the oligonucleotides/ribonucleotide sequence using during the ligation step, which is placed at the 5' end of a cDNA or RNA. This class II S restriction enzyme can then cleave the cDNA and produce the 5' end tag.

【0018】

Alternatively to biotin, a cap-binding protein (Pelletier et al. Mol Cell Biol 1995 15:3363-71) or an antibody that specifically binds to the cap structure can be used as the aforementioned selectively binding substance.

【0019】

Alternatively, one could use methods to attach oligonucleotides chemically to the cap structure as described by Genset. This method is based on the oxidation of cap (US patent 6,022,715). This allows (1) adding to the cap an oligonucleotides, which may contain the ClassIIs enzyme, (2) preparing first-strand cDNA synthesis which then switched second strand cDNA synthesis; after the second strand synthesis, the cDNA would be cleaved with Class II s enzymes to make a 5' tag, for subsequent formation of the concatamer.

【0020】

Alternatively, one could use the Use the cap-switch method as described by Clontech (US patent 5,962,272). One could prepare the first-strand cDNA in presence of a cap-switch oligonucleotide, which carries a recognition site for a substance capable to recognize nucleic acids and cleave them apart from the said recognition sequence such a Class II s restriction enzyme site. The cap switch mechanism let the first strand sequence continue on the cap-switch oligonucleotides. This can be followed by second cDNA strand, possibly also followed by PCR (as describes for instance in the SMARTTM Clontech cloning system), and finally it would be cleaved with the class II s to produce the 5' end TAGS.

【0021】

In another embodiment, when the quality of RNA allows it, one can prepare the cDNA by priming and extending the RNA until the cap-structure. Particular enzyme and reaction condition allow sometimes reaching the cap-site very efficiently (Carninci et al, Biotechniques, 2002). Even witho

ut a cap-selection it is possible to attach oligonucleotides in place of the cap structure, which carry Class II's restriction enzyme sites that would be later used to produce concatamers.

【0022】

The cap trapper method first synthesizes the first-strand cDNA with a reverse transcriptase by using RNA as a template. This can be conducted by a known method. The cDNA can be primed with an oligo-dT primer or, when the template RNA is mRNA, it can be primed with a random primer. It is advisable to add trehalose to the reactive solution because it raises the efficiency of reverse transcription reaction by stabilizing the reverse transcriptase. It is preferable to use 5-methyl-dCTP instead of standard dCTP, because it avoids internal cDNA cleavage with several restriction enzymes and prevents unintended cleavage with restriction enzymes to a considerable extent. In addition, after the first-strand cDNA synthesis, proteins and digested peptides might be removed by CTAB (cetyl trimethyl ammonium bromide) treatment, or other more general methods to purify cDNA.

【0023】

Next, a selectively binding substance is bound to the cap structure of mRNA. A "selectively binding substance" here means a substance that selectively binds to a specific substance, preferably but not limited to biotin. The cap structure is the structure at the 5' end of mRNA, which does not exist in transfer RNA (tRNA) or ribosomal RNA (rRNA). Therefore, even if total RNA was used as the starting substrate, the selectively binding substance only binds to mRNA. In addition, the selectively binding substance does not bind to mRNA if the cap structure at the 5' end has been cleaved. Biotin can be bound to the cap structure by a known method. For instance, the cap structure can be biotinylated by first oxidizing the diol groups on the cap structure by treating mRNA with an oxidizer

such as NaIO₄ and making then react with biotin hydrazide. Alternatively, any other methods known to a person trained in the state of the art of the preparation of full-length cDNAs can be utilized to selectively enrich 5'-ends according to the invention.

【0024】

Then, single-strand RNA is cleaved by means such as RNase I treatment. Any other RNase that can cleave single strand RNA but not cDNA/RNA hybrids or cocktails of RNases that can cleave the various single-strand RNA sequences at various specificity can be used alternatively. In an RNA/cDNA hybrid whose first-strand cDNA has not extended to the site corresponding to the 5' end site of RNA, the vicinity of the 5' end of RNA is single-stranded due to its failure to be hybridized with cDNA. Thus, the hybrid is cleaved at the single-stranded part and loses its cap structure through this step. Consequently, this step leaves only those mRNA/cDNA hybrids with cDNA that fully extends to the 5' end of mRNA to maintain the cap structure.

【0025】

A matching selectively binding substance fixed to a support, which selectively binds to the aforementioned selectively binding substance, is prepared. In the present specification, a "matching selectively binding substance" means a substance that selectively binds to the aforementioned selectively binding substance, which, in the case where the selectively binding substance is biotin, would be avidin, streptavidin or a derivative thereof that binds specifically to biotin or its derivatives. The support can favorably be, but is not limited to be, magnetic beads, particularly magnetic porous glass beads. Since magnetic porous glass beads to which streptavidin has been fixed are commercially available, such commercial streptavidin-fixed magnetic porous glass beads can be used favorably. Similarly other materials such as latex beads, latex magnetic bead

s, agarose beads, polystyrene beads, sepharose beads or alike could be used instead of porous glass beads. Furthermore, the invention is not limited to the use the biotin-avidin system but other binding substances could be used like a digoxigenin tag that would be attached to the cap structure and digoxigenin recognizing antibodies attached to a solid matrix.

【0026】

Following this, the aforementioned mRNA/cDNA hybrid with the cap structure is made to react with the aforementioned matching selectively binding substance fixed to the support in order to bind the selectively binding substance on the cap structure with the matching selectively binding substance on the support, thereby immobilizing the mRNA/cDNA hybrid with the cap structure on the support. When magnetic beads are used as the support, the magnetic beads can be quickly collected by applying a magnetic force. As mentioned above, the mRNA/cDNA hybrids that have the cap structure at this stage are only those with cDNA that fully extends to the 5' end of mRNA, so cDNAs containing a site complementary to the 5' end of mRNAs are selectively collected by this step, and Step 1 is completed. Meanwhile, in order to prevent non-specific binding to the support, it is preferable to treat the support with a large excess of DNA-free tRNA for blocking such binding before conducting this reaction. Other substances that are suitable for blocking the surface are nucleic acids or derivatives, for instance total RNA or oligonucleotides; proteins, for instance bovine serum albumine; polysaccharides, for instance glycogen, dextran sulphate, heparin or other polysaccharides. Alternatively, hybrid molecules containing parts of all of the above could be used to mask non-specific binding sites.

【0027】

The above focuses on the case where Step 1 is conducted by the cap tra

pper method, but other various methods can also be used as indicated as long as they can selectively collect cDNAs containing a site complementary to the 5' end site of mRNA.

【0028】

The following Step 2 selectively collects fragments containing a cDNA site that at least contains a site complementary to the 5' end site of mRNA.

【0029】

First, the first-strand cDNA that has been immobilized on the support is released. It can be conducted by treating the support with alkali, such as NaOH. Alternatively to alkali, an enzymatic reaction with RNaseH (which cleaves only the RNA hybridized to DNA) could be used. The alkali treatment releases the cDNA from the mRNA/cDNA hybrid, bound to the support through the cap on the mRNA and separates the cDNA from the mRNA to only leave first-strand cDNA on its own.

【0030】

Then, a linker carrying a sequence that can be recognized in a sequence-specific manner by a substance having an enzymatic activity that cleaves the recognized DNA outside the recognition sequence. An example of such substance is a Class IIIs restriction enzyme.

【0031】

In this embodiment, a linker that at least carries a Class IIIs restriction enzyme site, and a random oligomer part at the 3' end site, is ligated to the end of this first-strand cDNA, which corresponds to the 5' end of the aforementioned mRNA (i.e. the 3' end of the cDNA). For the later cloning of the 5' end sequence tags into concatemeres it is preferable but not essential to introduce a second recognition site into the linker, which should be distinct from the aforementioned recognition site used for the e.g. Class IIIs restriction enzyme.

【0032】

This can preferably be conducted as follows, by a method using a linker that carries a Class IIIs restriction enzyme site and a random oligomer part (SSLLM (single strand linker ligation method), Y. Shibata et al., BioTechniques, Vol. 30, No. 6, pp. 1250-1254, (2001)). The Class IIIs restriction enzyme is a restriction enzyme group that causes cleavage at parts other than the recognition site. An example includes but is not limited to the use of GsuI. GsuI treatment cleaves one of the strands at 16 bp downstream from the recognition site, and the other strand at 14 bp downstream from the recognition site. Another suitable example is MmeI, which cleaves respectively 20 and 18 bases apart its recognition sequence. The random oligomer part is located at the 3' end site of the linker, and though the number of bases is not particularly restricted, the recommended number is 5 to 9, or more preferably, 5 to 6. The Class IIIs restriction enzyme site should be located close to the aforementioned random oligomer part, so that the cleavage point comes within the cDNA, particularly relatively within the 5' side of the cDNA (i.e. the 3' side of the template mRNA). The linker should preferably be a linker for double-stranded DNA of which the aforementioned random oligomer part protrudes to the 3' side and provides the binding end. In addition, it is advisable to bind a selectively binding substance such as biotin to the linker in advance to facilitate its collection later.

【0033】

When the aforementioned first-strand cDNA is made to react with such a linker, the random oligomer part of the linker hybridizes with the 3' end site of the first-strand cDNA (i.e. the 5' end site of the template mRNA). Next, the second-strand cDNA is synthesized by using this linker as a primer and the first-strand cDNA as a template. This step can be con-

ducted by a standard method.

【0034】

Then, the obtained double-strand cDNA is treated with the above Class IIIs restriction enzyme. This step produces a double-strand cDNA fragment comprising a linker-derived part and a part derived from the 5' end site of the cDNA (the 5' end site of the second-strand cDNA). For instance, if *Gsu*I were to be used as the Class IIIs restriction enzyme and if there were to be a linker designed to locate the restriction site immediately upstream from the aforementioned random oligomer site, the obtained DNA fragment would include a site derived from the site on the 5' end side of the second-strand DNA (i.e. the site on the 5' end side of the mRNA) of the length of 16 bp (however, the complementary strand is 14 bp). In the case of the use of *Mme* I the length of the second-strand DNA fragment should increase to 20 and 18 bp respectively.

【0035】

Next, such DNA fragments are selectively collected. If a selectively binding substance (e.g. biotin) had been bound to the linker as above, the collection could be conducted similarly to Step 1 by using a support to which a matching selectively binding substance (e.g. streptavidin) would be fixed. This procedure completes Step 2, which selectively collects fragments containing a cDNA site, belonging to the first-strand cDNA, which at least contains a site complementary to the 5' end site of the aforementioned mRNA.

【0036】

The above explains the case where the SSLLM is used for Step 2, but Step 2 can also be carried out by any other method as long as the method can selectively collect fragments containing the 3' end site of the first-strand cDNA (the 5' end site of the template mRNA). For instance, it is possible to use exonuclease that cleaves the nucleotide in the 5'-3' di

rection at a controlled speed. The exonuclease treatment of the first-strand cDNA for a prescribed time period leaves a single-strand fragment comprising the 3' end site of the first-strand cDNA (the 5' end site of the template mRNA). It is possible to obtain only the targeted single-strand fragments by conducting treatment with a nuclease that only splits double-strand fragments. These fragments can be collected, joined with adaptors and cloned.

【0037】

The subsequent Step 3 forms concatamers by mutually ligating the collected fragments. Since there are multiple mRNAs and the linker hybridizes with the first-strand cDNA at the random oligomer part as above, the above method can obtain fragments containing multiple cDNAs derived from multiple mRNAs within a sample. Step 3 ligates these multiple fragments and forms concatamers. The ligation of the cDNA fragments can be carried out by a standard method, using commercial ligation kits. The ligation can be securely conducted but is not limited to a method which first is introducing a second linker providing a recognition site for a restriction enzyme that is distinct from the other recognition sites used at the earlier stages, which is then ligating two fragments into di-tags, and which is further ligating such ligated di-tag fragments into concatamers. The number of ligated fragments is not restricted, practically any number above two and preferably about 30. The obtained concatamers are preferably but not limited to be amplified or cloned by a standard method.

【0038】

The concatamers obtained in this way each comprise a site having the same base sequence (however, uracil in RNA would be thymine in DNA) as that of the 5' end site of the multiple mRNAs within the sample. Although it also comprises a part derived from the linker or linkers, the base sequence of the linker or linkers is already known, so the part derived fr

on the linker or linkers and the part derived from mRNA can be clearly distinguished by investigating the base sequence of the concatamer. Therefore, by determining the base sequence of the obtained concatamer, it is possible to find out the base sequences at the 5' end site of multiple mRNAs within the sample. The base sequences of a maximum of 16 or 20 bases at the 5' end site of each mRNA can be learned by the preferable mode of using *Gsu*I or *Mme* I. Information on 16 or 20 bases would be sufficient for almost definitely identifying the mRNA statistically and to judge whether or not it is a new mRNA. In addition, by determining the base sequence of the concatamer, it is possible to learn the base sequences at the 5' end site of mRNAs for the number of above fragments included in the concatamer (preferably 20 to 30), so information on the 5' end site of multiple mRNAs can be determined efficiently. The analysis of the concatamers can be automated by the use of computer software to distinguish between sequences derived from the 5'-ends and sequences derived from a linker or the linkers.

【0039】

When a new mRNA exists in a base sequence at the 5' end, the cDNA derived from the new mRNA can be obtained by conducting RT-PCR, making that site the forward primer and oligo-dT the reverse primer. It is also possible to amplify the mRNA by methods such as NASBA. Accordingly, the method of the present invention can be used for the cloning of new genes. Similarly, forward primers derived from 5'-end specific information can be used to amplify partial or full-length cDNA fragments from existing cDNA libraries.

【0040】

While the above method had used mRNA or total RNA within the sample as the starting substrate, Step 1 can be omitted by using an existing full-length cDNA library. In this way, information on the base sequences of

the 5' end site of multiple cDNAs (i.e. the 5' end site of the mRNAs used as templates for said cDNAs) contained in the full-length cDNA library can be efficiently obtained similarly to the above procedure.

【0041】

In some embodiments it could be desirable to obtain extended sequence information from the 5'-ends of transcribed regions. Such extended sequences may allow in specific cases for the identification of start sites of protein synthesis or a better mapping to genomic sequences. As described above the invention included in Step 2 the ligation of a linker to the 5' end of a cDNA. Such a linker can be modified by introducing a single-stranded overhang encompassing a sequence obtained from a concatamer to bind to and to be ligated to a specific nucleic acid fragment. After the ligation the linker can be used to enrich the DNA fragment by attaching the linker to a support from which it could be released after the enrichment. The linker can further be used as a primer to obtain extended sequence information on 5' ends.

【0042】

By investigating the base sequences of the concatamers or extended 5'-sequences obtained by the present invention, it is not only possible to clone new genes as described above, but also possible to investigate the expression profiles of genes within the sample. Furthermore, the technology can be used for various purposes such as to map transcription start sites in the genome, to map promoter usage patterns, for the analysis of SNPs in promoter regions, for creating gene networks by combining the expression analysis with information on promoters, alternative promoter usage and the other data, and for selective collection of the promoter site within fragmented genomic DNA. To select genomic fragments containing promoter sites, a fragment containing the same base sequence as the 5' end site of mRNA could be bounded to a support e.g. by using the aforem

entioned Biotin system, and hybridized to fragmented genomic DNA. Hybridized genomic DNA fragments could then be separated from a mixture of genomic fragments by using e.g. streptavidin-fixed magnetic beads, and cloned under standard conditions.

【0043】

Alternatively, one could avoid to make concatamers and use selected 5' end tags by ligating a mixture of full-length cDNAs to magnetic beads carrying homogeneous sequence of oligonucleotides, followed by ligation such as in the SSLLM, second strand cDNA preparation and cleavage with a Class IIIs restriction enzyme. The 5' end specific tag would be anchored specifically to the beads and would be used for the specific sequencing as done by Lynx therapeutics (US patents 6,352,828; 6,306,597; 6,280,935; 6,265,163; 5,695,934).

【0044】

For instance, oligonucleotides would have a "random part I", which will bind to 5' ends of cDNAs; and a code part of the oligonucleotide, which will be able to "tag" the ligation product. The oligonucleotide may be destroyed by exonuclease VII if not hybridized with a cDNA. The "decoder" oligonucleotides would be used to select out the sequence. The specific arrays of cDNAs on beads are then arrayed onto a solid surface, one per position, followed by parallel sequencing. If you look at 1 hole per 1 bead, you can make arrays of beads having specific oligonucleotides.

【0045】

By modifications as the aforementioned approaches for direct sequencing of 5' end the invention provides different means for the general analysis of 5' ends in the form of concatamers or the analysis of individual 5' ends, which were enriched by means of a 5' end specific selection.

【0046】

【実施例】

The present invention will now be described by way of examples thereof. It should be noted that the present invention is not restricted to the Examples. The experiments describe in the Examples can be performed by any person experienced in the state of the art of standard techniques in the field of Molecular Biology. Unless otherwise defined in the text, the technical terms, abbreviations, and solutions used in the Examples should have the same meaning as commonly understood by a person experienced to the state of the art in the field of the invention. A general description of such terms, abbreviations and solutions can be found in the common reagent section in Molecular Cloning (Sambrook and Russel, 2001). All publications mentioned herein are incorporated into this document by reference to be disclosed and to describe the methods and/or materials herein.

【0047】

Example 1:

Preparation of total RNA from tissue

In the literature a variety of different approaches for the preparation of RNA have been described, which are known to a person experienced in the state of the art. All such approaches should allow the preparation of a plurality of RNA samples derived from biological materials including tissues and cells, which are suitable for the invention. Below two such procedures are described in detail.

Buffers and solutions:

- a) Solution D: 4M guanidinium thiocyanate, 25mM sodium citrate (pH7.0), 100mM 2-mercaptoethanol and 0.5% n-lauryl-sarcosine.
- b) RNase-free CTAB/UREA solution: 1% CTAB (Sigma), 4M UREA, 50mM Tris-HCl (PH 7.0), 1mM EDTA (pH 8.0).
- c) Water equilibrated phenol as described in Molecular Cloni

ng (Sambrook and Russel, 2001).

Phosphate-buffer saline (PBS) as described in Molecular Cloning (Sambrook and Russel, 2001)

5 M Sodium chloride

7 M Guanidium choride

Rnase free dd-water

【0048】

Protocol for total RNA preparation

Dissect the tissue as fast as possible in a cooled dish.

Roughly evaluate the volume of tissue in a 50 ml falcon tube. The best quantity of tissue is between 0.5-1 g of tissue for 20 ml Solution D

Add 2 ml of 2M sodium acetate (pH 4.0) and 16 ml of water-equilibrated phenol.

Mix by a vortex. Add 4 ml of chloroform and shake vigorously by your hands and a vortex. Let it stay on ice for 15 min.

Centrifuge it at 6,000 rpm for 30 min at 4 °C

Transfer the upper aqueous phase to new tube by pipetting (25 ml) and recover approximately 20 ml thereof.

Precipitate the RNA from the aqueous phase by adding 1 equal volume of Isopropanol (in this case, approximately 20 ml), store on ice for 1 h.

Centrifuge at 7,500 rpm for 15 min at 4 °C: RNA is pelleted by centrifugation.

The pellet is washed twice with 70% ethanol, each time followed by centrifugation at 7,500 rpm for 2 min, in order to remove the SCN salts.

CTAB removal of polysaccharides. Selective CTAB precipitation of mRNA is performed after complete RNA re-suspension in 4 ml of water. Subsequently, 1.3 ml of 5 M NaCl is added and the RNA is then selectively precipitated by adding 16 ml of a CTAB/urea solution.

Centrifuge for 15 min at 7500 rpm (9500 x g), discard the aqueous phase.

Resuspend the RNA pellet in 4 ml of 7 M Guanidinium Chloride.

Re-suspended RNA is finally precipitated by adding 8 ml of ethanol. Incubate on -20°C for 1-2 hours (or longer) and centrifuge for 15 min at 7,500rpm, 4°C. At the end, wash the pellet with 5 ml of 70% ethanol.

Centrifuge again at 7,500 rpm for 5 min.

Discard the supernatant.

Re-suspend RNA in 500-1000 microL of RNase-free dd-water.

【0049】

Preparation of a mRNA fraction from total RNA

The mRNA fraction of total RNA preparations can be isolated by the use of commercial kits such as the MACS mRNA isolation kit (Milteny) or polyA-quick (Stratagene), which provide satisfactory yield of mRNA under the recommended conditions. One cycle of oligo-dT selection of the mRNA is sufficient. It is advisable to redissolve the poly-A⁺ RNA at a high concentration of 1 to 2 microG/microL.

【0050】

Preparation of a plurality of RNA samples from a cDNA library

Alternatively, a plurality of nucleic acids corresponding to the 5' ends of genes can be obtained from existing cDNA libraries, which were cloned into expression vectors. By standard methods known to a person familiar with the state of the art of molecular biology approaches, from such libraries RNA transcripts can be obtained by in vitro transcription reactions using e.g. a T3, T7 or SP6 RNA polymerase. Such an approach can be performed by first linearization of the plasmid DNA with appropriate restriction endonucleases. The restriction enzyme can be chosen to allow for the transcription of the sense RNA. In the case of libraries obtained in the vector pFLC III (Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraiki T, Hirozane T, Watahiki A, Shibata K, Konno H, Muramatsu M, Hayashizaki Y.) Balanced-size and long-size cloning of full-length, cap-trapped

cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis, Genomics, 2001 Sep;77(1-2):79-90), the vector can be linearized by cleavage with one of the homing endonucleases I-Ceu I or PI-Sce I to avoid a truncation of the inserts. For the digest mix in a tube

Plasmid DNA 100 microG

10x buffer 40 microL

Restriction enzyme 100 u

ddH₂O ad 400 microL

Incubate at appropriate temperature for at least 2h and analyze 1 microL of the reaction mixture by agarose gel electrophoreses. If the digest is completed, add:

0.5 M EDTA 8 microL

10% SDS 8 microL

Proteinase K (10 mg/ml) 5 microL

Incubate for 15 min at 45°C before extracting sample with 500 microL phenol/chloroform. The aqueous phase is to be re-extracted twice with 500 microL chloroform. Finally linearized DNA is precipitated with isopropanol or ethanol under standard conditions and dissolved in 50 microL TE.

【0051】

In vitro RNA synthesis:

Mix in a tube under RNase free conditions:

Linearized plasmid DNA 20 microG

5x T7 or T3 buffer 200 microL

0.1 M DTT 100 microL

2 mg/ml BSA 40 microL

10 mM rNTPs 50 microL

T7 or T3 RNA polymerase 10 microL

ddH₂O ad 1000 microL

Incubate at 37°C for 3 to 4 h before adding:

10 mM Calcium Chloride 10 microL

1U/microL DNase RQ1 5 microL

Incubate at 37°C for 20 min before adding:

0.5 M EDTA 10 microL

10 mg/ml Protease K 5 microL

Incubate at 45°C for 30 min, before addition of Sodium Chloride to a final concentration of 1M. Phenol/Chloroform extraction followed by re-extraction with Chloroform should be performed under standard conditions, and the RNA transcripts can be finally collected by Isopropanol or Ethanol precipitation. The pellet is to be resuspended in 200 microL of water or TE. The quality of the RNA transcripts should be confirmed by agarose gel electrophoresis and quantification.

【0052】

2. : First strand cDNA synthesis

Buffers and solutions

Saturated Trehalose, about 80% in water (crystals will remain), low metal content

4.9 M high purity sorbitol

Optionally: Takara GC-Taq buffer

【0053】

Enzymes and buffers

RNase H⁻ reverse transcriptase Superscript II (Invitrogen) and buffer or other reverse transcriptases.

【0054】

Nucleic acids and oligonucleotides

Purified, first-strand oligo-dT primer (Sequence for primer used: 5'-GAGAGAGAGAGGATCCTCTGGAGAGTTTTTTTTTTVN-3'). Alternatively or additionally, random primer (dN₆-dN₉), where N is any nucleotide.

mRNA, recommended 2.5 to 25 microG or alternatively, total RNA, 5-50 microG

【0055】

Radioactive compounds

[alpha-³²P] dGTP

【0056】

Protocol A: Trehalose-Sorbitol enhanced

To prepare the 1st strand cDNA, put together the following reagents in three different 0.5 ml PCR tubes (A, B, and C)

【0057】

Tube A: in a final volume of 21.3 microL, add the following:

mRNA 2.5-25 microG

or total RNA, 5-50 microG

1st strand primer (2 microG/microL) 14 microG (7 microL)

Total volume: 22 microL

Heat the mixture (mRNA, primer) at 65°C for 10 min to dissolve the secondary structures of mRNA.

Tube B: in a final volume of 76 microL, add the following:

5X 1st strand buffer 28.6 microL

0.1 M DTT 11 microL

dATP, dTTP, dGTP, and 5-methyl-dCTP 10 mM each 9.3 microL

4.9 M sorbitol 55.4 microL

Saturated trehalose 23.2 microL

RNase H- Superscript II reverse transcriptase (200 U/microL)

15.0 microL

Final volume: 142.5 microL

【0058】

Prepare a cycle (on a thermal cycle) with: 40°C, 4 min; 50°C, 2 min; 56°C, 60 min.

If total RNA is used as the starting material, prepare a cycle with:

40°C, 2 min, -0.1°C/sec to 35°C; 50°C, 2 min; 56°C, 60 min.

Alternatively: prime the cDNA with a random primer (dN₉, N= any nucleotide) at 25°C.

【0059】

Tube C:

1~1.5 microL of [alpha-³²P] dGTP.

【0060】

For a cold-start operate as follows:

Quickly mix tubes A and B on ice.

Transfer in tube C 40 microL of the A+B mixture.

Tubes A+B and C should be quickly transferred immediately at 40°C of the step 1 of the above cycling program to anneal at 40°C for 4 minutes.

Let the reaction proceed following the thermal cycler setting.

【0061】

For a hot-start, operate as follows:

Transfer the tubes A, B, C on the thermal cycler

Start the cycling

When the temperature reaches 42°C, quickly mix tubes A and B.

Transfer in tube C 40 microL of the A+B mixture.

Let the reaction proceed following the thermal cycler setting.

【0062】

Protocol B: GCI-Trehalose-Sorbitol enhanced

Tube A: in a final volume of 22 microL, add the following:

mRNA 5~25 microG

(precipitate with ethanol and re-suspend directly with the primer)

or total RNA, up to 50 microG (for the small-scale protocol)

Purified 1st strand cDNA primer (2 microG/microL) 14 microG (7 microL)

Final volume: 22 microL

Tube B: add the following:

2 X GC I (LA Taq) buffer (TaKaRa)	75 microL
dATP, dTTP, dGTP, and 5-methyl-dCTP, 10 mM each	4 microL
4.9 M sorbitol	20 microL
Saturated trehalose (approximately 80%)	10 microL
Superscript II reverse transcriptase (200 U/microL)	15 microL
ddH ₂ O	4 microL
Final volume:	128 microL
Tube C:	
alpha- ³² P-dGTP	1.5 microL

For the rest of the procedure, follow exactly the point as in the normal reaction condition. Prepare (in advance) a thermal cycler with the following cycle:

42°C, 30 min; 50°C, 10 min; 55°C, 10 min; 4°C, indefinite time.

【0063】

Operate as follows:

- 1) Transfer the tubes A, B, C on the thermal cycler
- 2) Start the cycling
- 3) When the temperature reaches 42°C, quickly mix tubes A and B.
- 4) Transfer in tube C 40 microL of the A+B mixture.
- 5) Let the reaction proceed following the thermal cycle setting.

At the end, stop the reaction with EDTA at 10 mM final concentration.

Then incorporation of [alpha³²P]GTP is measured and the yield of cDNA is calculated. Calculation of the amount of cDNA by measuring [alpha³²P]GTP is useful for monitoring whether the processes are accurately proceeding or not.

【0064】

3. CTAB precipitation of the first-strand cDNA

Buffers and solutions

CTAB solution as described in Example 1

After measuring the radioactivity, transfer both the "hot" and "cold" 1st strand synthesis (tube B and C) to a tube and perform CTAB precipitation as follows.

Mix the tube B and C from the first strand; to the mixture add:

3 microL of 0.5 M EDTA (final concentration of 10 mM)

2 microL of 10 microG/microL Proteinase K.

Incubate at 45°C or 50°C for at least 15 min, and as long as 1 hour.

To the 128-142 microL volume of the first strand cDNA reaction, add:

32 microL of 5 M Sodium Chloride (RNase free)

320 microL of CTAB-Urea solution

Incubate at room temperature for 10 min.

Centrifuge at 15,000 rpm for 10 min

Remove supernatant.

Carefully re-suspend with 100 microL of 7M Guanidinium Chloride

Add 250 microL of ethanol and leave on ice or 20 to -80°C for 30-60 min

Centrifuge at 15,000 for 10 min. Remove the supernatant.

Subsequently, wash the pellet twice with 800 microL of 80% ethanol. Each time, add 80% ethanol to the tube and centrifuge for 3 min. at 15,000 rpm.

Re-suspend cDNA in water 46 microL.

【0065】

4. Cap-trapping, oxidation and biotinylation of the cap

Buffers and solutions

1 M sodium acetate buffer, pH 4.5

1M citrate buffer, pH 6.0

NaIO₄, solution >100 mM.

SDS 10%

Biotinylation buffer: 33 mM Sodium citrate, pH 6.0, and 0.33% SDS.

10 mM Biotin Hydrazide long arm (MW = 371.51; 3.71 mg/ml = 10 mM) in citrate/SDS buffer.

Cap biotinylation: (A) Oxidation of the diol groups of mRNA

In a final volume of 50 to 55 microL, add the following:

The re-suspended cDNA sample

3.3 microL of 1 M sodium acetate buffer, pH 4.5

A freshly prepared solution of NaIO₄ to a final concentration of 10 mM

Incubate on ice in the dark for 45 min.

Finally, precipitate the cDNA:

To simplify the downstream process, add 1 microL of glycerol 80%.

Vortex.

Add 0.5 microL of 10% SDS, 11 microL of 5 M sodium chloride and 61 microL of isopropanol.

Incubate at 20 or -80°C for 30 min in the dark.

Centrifuge for 15 min at 15,000 rpm.

Remove supernatant.

Add 500 microL of 80% ethanol

Centrifuge at 15,000 rpm for 2-3 min.

Discard the supernatant

Repeat steps 12-13

Re-suspend the cDNA in 50 microL of water.

Biotinylation: (B) Derivatization of the oxidized diol groups

To the cDNA (50 microL), add 160 microL of the dissolved biotin hydrazide long arm in the reaction buffer. Perform the reaction in 210 microL (final volume).

Incubate overnight (10-16 hours) at room temperature (22-26°C).

Subsequently, to precipitate the biotinylated cDNA, add:

75 microL 1 M Sodium citrate, pH 6.1

5 microL of 5 M Sodium chloride

750 microL of absolute ethanol

Incubate on ice for 1 hour or at 80 or -20°C for 30 min or longer.

Centrifuge the sample at 15,000 rpm for 10 min

Wash the precipitate twice with 70% or 80% ethanol and centrifuge.

Discard the supernatant and repeat the wash. Dissolve the cDNA in 175 microL of TE (1 mM Tris, pH 7.5, 0.1 mM EDTA).

Cap-trapping and releasing the 5' ends of cDNA Enzymes and buffers

RNase ONE (Promega) and its reaction buffer

To the cDNA sample add, in a final volume of 200 microL:

20 microL of RNase I buffer (Promega).

1 units of RNase I (Promega, 5 or 10 U/microL) per each 1 microG of starting mRNA or total RNA (in case of small scale protocol) used for first strand cDNA synthesis.

Incubate at 37°C for 30 min.

To stop the reaction, put the sample on ice and add

4 microL 10% SDS and

3 microL of 10 microG/microL Proteinase K.

Incubate at 45°C for 15 min.

Extract once with 1:1 Tris-equilibrated phenol:chloroform, then load the aqueous phase into Microcon -100.

Perform a back extraction with water and load again into the Microcon-Centricon 100 filter.

Perform one round of Microcon separation

8-b) Dissolve completely the pellet with 20 microL of 0.1 x TE

【0066】

Magnetic beads blocking

Materials

Streptavidin-coated MPG (CPG inc., New Jersey)

Buffers and solutions

Binding buffer: 4.5 M NaCl, 50 mM EDTA, pH 8.0

Special equipments

A magnetic stand to hold 1.5 ml tubes is required.

To further minimize the non-specific binding of nucleic acids, magnetic beads are pre-incubated with DNA-free tRNA (10mg/ml).

For each preparation, pre-incubate 500 microL of magnetic beads (per 25 microG of starting mRNA) with 100 microG of tRNA.

Incubate on ice for 30 min with occasional mixing.

Separate the beads with a magnetic stand (for 3 min) and remove the supernatant.

Wash for 3 times with 500 microL of binding buffer

【0067】

5'-ends cDNA capture and release

To capture the full-length cDNA, mix the RNaseI-treated cDNA and wash beads as follows:

- 1) Re-suspend the beads in 500 microL of wash/binding buffer.
- 2) Transfer 350 microL of the beads into the tube containing the biotinylated first-strand cDNA.
- 3) After mixing gently rotate the tube for 10 min at 50

°C,

4) Transfer 150 microL of the beads into the tube containing the biotinylated first-strand cDNA and 350 microL of beads.

5) After mixing gently rotate the tube for 20 min at 50 °C.

Separate the beads from the supernatant on a magnetic stand.

Washing the beads

Gently wash the beads with 0.5 ml of the indicated buffer to remove the nonspecifically absorbed cDNAs.

2 x with washing/binding solution.

1 x with 0.3 M NaCl/ 1mM EDTA

2 x with 0.4% SDS/ 0.5 M NaOAc/ 20 mM Tris-HCl pH 8.5/ 1mM EDTA.

2 x with 0.5 M NaOAc/ 10 mM Tris-HCl pH 8.5/ 1mM EDTA.

Alkali release (see below)

Alkali full-length cDNA release from beads

Add 100 microL of 50 mM NaOH, 5 mM EDTA.

Briefly stir and incubate 5 min at RT with occasional mixing.

Separate the magnetic beads and transfer the eluted cDNA on ice.

Repeat the elution cycle with 100 microL of 50 mM NaOH, 5 mM EDTA, two more times until most of the cDNA, 80-90% as measured by monitoring the radioactivity, can be recovered from the beads.

Adding a 5'-end primable site to the cDNA

RNase step

Enzymes and buffers

- RNase ONE™ and its buffer (Promega)

Add 50 microL of 1 M Tris-HCl, pH 7.0 in tubes on ice and mix quickly.

Add 1 microL of RNase I (10U/microL) and mix quickly.

Incubate at 37 °C for 10 min.

To remove the RNaseI, treat the cDNA with Proteinase K and phenol/chloro

form extraction including back extraction.

Add 3 microG of glycogen. Treat the cDNA with one cycle of Microcon-100.

Fractionation of cDNA before adding a primable site

Materials

Amersham-Pharmacia S-400 spun kit or alternative kits

Buffers and solutions

Column buffer: 10 mM Tris, pH 8.0, 1 mM EDTA, 0.1 % SDS, and 100 mM NaCl

Column buffer without SDS: 10 mM Tris, pH 8.0, 1 mM EDTA and 100 mM NaCl

S-400 spun column chromatography

Detailed protocols are described in the kits. This is the running protocol of S-400 spun columns.

Shake the column

Break the seal and transfer in a 2 ml tube

Centrifuge at 3,000 rpm 1 min (+ 4 °C)

Add the cDNA (< 20 microL volume)

After cDNA, add 80 microL of water

Centrifuge 2 min at 3000 rpm

Concentrate by Microcon 100 or precipitate with isopropanol. Recovery should exceed 80%.

【0068】

6. SSLLM

Materials

S-300 spun column chromatography kit (Amersham-Pharmacia)

Buffers and solutions

Column buffer: 10mM TrisHCl pH 8.0, 1mM EDTA, 0.1% SDS, 100mM NaCl.

Enzymes and buffers

Takara DNA Ligase KIT II.

Nucleic acids and oligonucleotides

In the Example given here, the recognition sites for the restriction enzymes Bgl II, Gsu I and Mme I are introduced, however, the invention is not dependent or limited to the use of those restriction enzymes and their recognition sites. In particular, Bgl II (recognition site: AGATCT) can be replaced by any endonuclease suitable for cloning. Other example for such enzyme could include Asc I (recognition site: GGCGCGCC) or Xba I (recognition site: TCTAGA).

- Synthesize the following oligonucleotides containing the Gsu I restriction site.

Oligonucleotide Bg-Gsu-GN5:

5' -Biotin-AGAGAGAGAACTAGGCTTAATAGGTGACTAGATCTGGAGGNNNN-3' ;

Oligonucleotide Bg-Gsu-N6:

5' -Biotin-AGAGAGAGAACTAGGCTTAATAGGTGACTAGATCTGGAGNNNNNN-3' ;

Oligonucleotide Bg-Gsu-down:

5' P-CTGGAGATCTAGTCACCTATTAAGCCTAGTTCTCTCT-NH₂ 3' .

- Synthesize the following oligonucleotides containing the Mme I restriction site.

Oligonucleotide Bg-Mme-GN5:

5' -Biotin-AGAGAGAGAACTAGGCTTAATAGGTGACTAGATCTTCCRACGNNNNN-3' ;

Oligonucleotide Bg-Mme-N6:

5' -Biotin-AGAGAGAGAACTAGGCTTAATAGGTGACTAGATCTTCCRACNNNNNN-3' ; Oligonucleotide Bg-Mme-down:

5' P-GTYGGAGATCTAGTCACCTATTAAGCCTAGTTCTCTCT-NH₂ 3' .

Where R stands for G or A and Y stands for C or T.

P means that the oligonucleotide must be 5' phosphorylated and NH₂ indicates that an amino-group is added to avoid non-specific ligation and possible hairpin priming.

Oligonucleotides should be purified by acrylamide gel electrophoresis following standard techniques as the first-strand cDNA primer with 10% acr

ylamide electrophoresis (Sambrook and Russel, 2001). Oligonulceotides should be extracted with phenol/chloroform, chloroform and precipitation with 2 volumes of ethanol as for the first-strand cDNA primer.

【0069】

Preparation of the linkers.

After OD checking and mixing Bg-Gsu-GN5, Bg-Gsu-N6 and "down" oligonucleotides at ratio 4:1:5, at least 2 microG/microL of DNA; add NaCl at 100 mM final concentration. The oligonulceotides are annealed at 65°C for 5min, 45°C for 5min, 37°C for 10min, 25°C for 10min.

【0070】

Ligation of the first-strand cDNA

Use 2 microG of linker mixture for up to 1 microG single-strand cDNA. Mix linkers and cDNA (final volume: 5 microL)

Heat at 65°C for 5min to melt secondary structures of single-strand cDNA
Transfer the linker and cDNA mix on ice.

Add 5 microL of the solution II from the TAKARA DNA ligation Kit.

Add 10 microL of solution I of the kit.

Incubate at 10°C overnight (at least >10 hours).

At the end of the ligation reaction, stop the reaction by adding 1microL of 0.5 M EDTA, 1 microL of 10% SDS, 1microL of 10 mg/ml Proteinase K, 10 microL of water, and incubate at 45°C for 15 min.

Treat with phenol/chloroform, chloroform and back extract (see appendix) with 60 microL of column buffer

After the ligation, remove the excess linker with S-300 spin column chromatography

- 1) Shake the column several times and then let it stand upright.
- 2) Remove the upper cap, then the bottom one.
- 3) Drain the buffer of the column. Apply 2 ml of the column buffer and drain twice by gravity.

Put the column into a 15 ml centrifuge tube, then centrifuge at 400 x g for 2 min in a swing-out rotor at room temperature.

Apply 100 microL of buffer to the column, then centrifuge at 400 x g for 2 min. Check the eluted volume. If it is different from the input (100 microL), repeat this step until the eluted volume is the same as the added one.

Set a 1.5 ml tube, after cutting off the cap, into the 15 ml centrifuge tube, and then apply the sample into the column. Centrifuge at 400 x g for 2 min.

Collect the eluted fraction in a separate tube. Apply to the column 50 microL of buffer, repeat the centrifugation and collect the fraction in a separate tube.

Repeat step 6 for 3 to 5 more times; keep the eluted fractions separate. Collected fractions should be counted in a scintillation counter. Usually mix the first 2-3 fractions (80% of cpm of cDNA).

Add NaCl to a final concentration of 0.2 M, precipitated the cDNA by adding equivalent of isopropanol.

After precipitation and washing twice with 80% cold ethanol, re-suspend with water.

Second-strand cDNA

Setting the 2nd strand cDNA program on the thermal cycler as follows:

Step 1	5 min at 65 °C
Step 2	30 min at 68 °C
Step 3	72 °C for 10 min
Step 4	+4°C

Procedure for the second strand cDNA

Second strand steps, mix in a test tube:

The cDNA

6 microL of LA-Taq polymerase buffer (Takara)

6 microL of 2.5 mM (each) dNTP's (Takara)

0.5 microL of [α -32P] dGTP (optional to follow the incorporation)

After starting the 2nd strand program, put the tube on the thermal cycle
r.

Add to tube 3 microL of 5 U/microL of LA Polymerase or alternative therm
ostable polymerase cocktails, when the samples are at 65°C, during the fi
rst step.

Mix quickly but thoroughly

At the end of the cycle of the thermal cycler, stop the reaction by addy
ing 10 mM EDTA (final concentration) and clean up the reaction by Protei
nase K treatment, Phenol-chloroform extraction and ethanol precipitation
(see Sambrook and Russel, 2001, Molecular Cloning, CSHL press, NY).

【0071】

11. Cleavage of cDNA

The cDNA should then be cleaved with the Class IIIs restriction enzyme li
ke Gsu I given in this Example.

Buffer (10X) (MBI Fermentas) 10 microL

GsuI(1U/microL) (use 5U/microG DNA) Y microL

ddH₂O X microL

Final volume 100 microL

Where the Y and X vary depending on the quantity of cDNA

1) Incubate at 37°C for 1 hour.

2) Added 0.5M EDTA 2 microL.

3) Incubated at 65°C for 15 min. to inactivate the enzyme

Prepare the magnetic beads

Prepare the appropriate quantity of CPG-MPG (Magnetic porous glass beads

). The same considerations made for the cap-trapper step are valid at this point.

Prepare 200 microl of GPG- beads.

Add 5 microG of tRNA (20 mg/ml).

Incubate at RT for 10-20 min or on ice for 30-60 min, with occasional shaking

Transfer the beads on a magnetic stand for 3 minutes and remove the aqueous phase.

Wash 3 times with: 1M NaCl, 10 mM EDTA use at least a volume equivalent to the starting volume of beads.

Re-suspend beads in 1M NaCl, 10 mM EDTA equivalent to the starting volume of beads.

【0072】

7. Released cDNA tags

Mixed washed beads and GsulI cut sample.

Incubate at RT for 15 min with occasional gentle mixing

Let it stand on magnetic rack for 3 min.

Recover the supernatant.

Rinse 4X with 500 microl of 1X B&W buffer (binding and washing buffer= 5 mM Tris, pH 7.5, 0.5 mM EDTA, and 1 M NaCl) containing 1X BSA (bovine serum albumin) wash.

Wash 2X with 200 microl of 1X ligase buffer (NEB).

【0073】

8. Ligating linkers to bound cDNA: II linker ligation.

In this Example a linker with a recognition site for the restriction enzyme Eco RI is used. However, the invention is not dependent or limited to the use of Eco RI in the second linker. Any other restriction enzyme and its recognition site can be used depending on their convenience for cloning the concatamers.

Oligonucleotides to be synthesized:

5'-GAGAGAGAGACTTAGGTGACACTATAGAAGAGTCCTGAGAATTCNN-3'

5'-P-GAATTCTCAGGACTCTTCTATAGTGTACACCTAAAGTCTCTCTC-3'

The oligonucleotides are purified and annealed as described for the Linker 1.

LoTE (1 mM Tris, pH 7.5, and 0.1 mM EDTA) 20 microL suspended and add linker II (0.4 microG/microL)

Heat the tube at 65 °C for 5min, then let sit at room temperature for 15 min.

Add TaKaRa ligation kit II solution II 25microL and solution I 50microL. Incubated at 16 °C overnight.

After ligation, wash 4 times with 500 microL 1X B&W buffer containing 1X BSA.

Wash once with 200 microL 1X B&W buffer and twice with 200 microL 1XBglI I buffer containing 1X BSA.

【0074】

Release of cDNA tags using the Tagging Enzyme

Add to the sample the following

-	LoTE	X microL
-	10X buffer	10 microL
-	Bgl II	Y microL

Make up the volume to a total of 100 microL.

- 1) Incubate at 37°C for 1 hour, gently mixing intermittently.
- 2) Place on magnet, collect supernatant into new tube. The supernatant contains the released 5' end fragments.

3) Raise volume to 200 microL with LoTE.

To 200 microL of sample (the 5' ends, tagged with linkers) add:

133 microL 7.5M NH4OAc

3 microL 1microG/microL glycogen

340 microL Isopropanol

Incubate at 20 or -80°C for at least 30 min.

Spin for 20min at 4°C at 15,000 rpm in a micro-centrifuge. Remove the supernatant. Wash the pellet twice with 80% or 70% ethanol. Centrifuge for 3 min at 15,000 rpm and removed the ethanol wash. At the end, re-suspended in 10 microL LoTE.

【0075】

Ligating tags to form di-tags

The 5' ends of cDNAs are ligated to form di-tags.

1) Add the TaKaRa ligation Kit II solution II 10 microL and solution I 20 microL.

2) Incubate overnight 16°C.

3) Added 10 microL of ddH₂O, 1 microL of 0.5M EDTA, 1 microL of 10% SDS 1 and 1 microL of 10 microG/microL Proteinase K.

4) Incubate at 45°C for 15min.

5) Extract once with 1:1 Tris-equilibrated phenol:chloroform aqueous phase. After phenol-chloroform and chloroform, and back extraction.

6) Removal the smallest cDNA fragment with a G-50 spun-column (Size exclusion).

7) precipitate with isopropanol by adding 5 microG of glycogen as carrier.

100 microL sample

67 microL 7.5M NH₄OAc

5 microL glycogen

180 microL Isopropanol

8) Spin for 20 min at 4 °C.

9) Wash twice with 80% or 70% ethanol, centrifuge and remove the ethanol.

【0076】

12. Cleavage of cDNA with anchoring enzyme

1) Re-suspend the sample in 5 microL of LoTE. Add then in order:

LoTE X microL

10X EcoRI restriction buffer 5 microL

EcoRI Y microL (use 20 Units of EcoRI)

Bring up the volume to a total of 50 microL.

2) Incubate at 37°C for 1hour.

3) Add 1 microL of 0.5M EDTA, 1microL of 10% SDS 1 and 1 microL of 10 microG/microL Proteinase K 10%.

4) Incubate at 45 °C for 15min.

5) Extract once with 1:1 Tris-equilibrated phenol:chloroform aqueous phase. After phenol-chloroform and chloroform, and back extraction

6) precipitate with isopropanol by adding 5 microG of glycogen as carrier.

100 microL sample

67 microL 7.5M NH40Ac

5 microL glycogen

180 microL Isopropanol

8) Spin for 20 min at 4°C.

9) Wash twice with 80% or 70% ethanol, centrifuge and removed the ethanol wash each time.

【0077】

11. Ligation of di-tags to form concatemers
- 1) Resuspended LoTE 5 microL.
- 2) Added TaKaRa ligation kit II solution II 5 microL and solution II 10 microL.
- 3) Incubate 1.5 hours at 16 °C.
- 4) Added 0.5M EDTA 1 microL, 10% SDS 1 microL, 10 microG/microL Proteinase K 1 microL.
- 5) Incubate at 45°C for 15min.
- 6) Extract once with 1:1 Tris-equilibrated phenol:chloroform aqueous phase. After phenol-chloroform and chloroform, and back extraction.
- 7) precipitate with isopropanol by adding 5 microG of glycogen as carrier.

100 microL sample

67 microL 7.5M NH4OAc

5 microL glycogen

180 microL Isopropanol

- 8) Spin for 20min at 4°C.

- 9) Wash twice with 80% or 70% ethanol, centrifuge and removed

Resolved 5 microL ddH₂O.

【0078】

Example 2:

The above-obtained concatamers are to be further ligated into a cloning vector such as pBlueascript II KS+ (Stratagene). A large variety of cloning vectors are known in the field, which can be used for invention.

Standard Ligation:

Mix a three time excess of concatamer DNA and 100 ng of an appropriate vector linearized with Eco RI in a volume of 5 microL. Then mix 5 microL

of Solution I of DNA Ligation Kit Ver.2 (Takara) to the insert/vector mixture. Incubate the tube at 16°C for 12-16 h.

【0079】

Transformation:

To remove salt from the ligation solution, precipitate DNA after the addition of 2 microG of Glycogen (Roche), 20mM Sodium Chloride and 80% ethanol. The DNA pellet is washed twice with 150 microL of 80% of ethanol, and the pellet is then dissolved in 10 microL of water. Using 1 microL of desalted ligation solution, ElectroMAX™ DH10B™ Cells (Invitrogen) are transformed using Cell-Porator or alike (Biometrer according to the transformation procedures described in the manufacturer's manual. Transformed bacteria are plated on a selective medium and grown overnight. Positive clones are to be isolated from those plates for further characterization of the concatamers.

【0080】

Example 3: Sequencing of concatamers

Sequencing of concatamers is performed using primers nested in the flanking regions of the cloning vector and a BigDye Terminator Cycle Sequencing Ready Reaction Kit v2.0 (Applied Biosystems) and an ABI3700 (Applied Biosystems) sequencer according to the manufacturer's product descriptions. The concatamers are sequenced from both ends to cover their entire sequence.

【0081】

Example 4: Identification of 5'-end sequence tags

The sequences obtained from Concatamers are characterized by the structure of the di-tags as presented in Figure 5. Defined regions holding the recognition sites for the restriction enzymes used during the cloning steps flank each 5' end specific sequence tag. Therefore the 5' end specific sequence tags can be identified by a manual sequence analysis or by

an automated process using an appropriate computer program. Individual 5' end specific sequence tags can be stored in a computer file or a data base system.

【0082】

Example 5: Characterization of 5'-end sequence tags

5' end specific sequence tags can be analyzed for their identity by standard software solutions to perform sequence alignments like NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>), FASTA, available in the Genetics Computer Group (GCG) package from Accelrys Inc. (<http://www.accelrys.com>) or alike. Such software solutions allow for an alignment of 5' end specific sequence tags among one another to identify unique or non-redundant tags, which can be further used in

Database searches

Building a 5'-end sequence database

Gene identification using a 5'-end sequence database

An example of a BLAST search in GenBank using a 5' end specific tag is given below: The 16 bp tag (5'-ACC TCC CTC CGC GGA G) is derived from the 5' end of Human TGF- β 1: JBC 264 (1989) 402-408.

Query= (16 letters) (ACCTCCCTCCGCGGAG)

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)

1,205,903 sequences; 5,297,768,116 total letters

Score E

Sequences producing significant alignments: (bits)

Value

gi|10863872|ref|NM_000660.1| Homo sapiens transforming grow... 32

1.1

gi 18590091 ref XM_085882.1 Homo sapiens similar to transf...	32
1.1	
gi 11424057 ref XM_008912.1 Homo sapiens transforming grow...	32
1.1	
gi 7684381 gb AC011462.4 AC011462 Homo sapiens chromosome 1...	32
1.1	
gi 15027087 emb AL389894.4 LMFLCHR4A Leishmania major Fried...	32
1.1	
gi 1943914 gb U70540.1 LMU70540 Leishmania mexicana amazone...	32
1.1	
gi 37097 emb X05839.1 HSTGFBG1 Human transforming growth fa...	32
1.1	
gi 37092 emb X02812.1 HSTGFB1 Human mRNA for transforming g...	32
1.1	
gi 340526 gb J04431.1 HUMTGFB1PR Homo sapiens transforming ...	32
1.1	
gi 18858696 ref NM_131728.1 Danio rerio forkhead box Cla (...	30
4.2	
gi 12004937 gb AF219949.1 AF219949 Danio rerio forkhead tra...	30
4.2	
gi 193604 gb M13366.1 MUSGPDX Mouse glycerophosphate dehydr...	30
4.2	
gi 193601 gb M25558.1 MUSGPD Mouse glycerol-3-phosphate deh...	30
4.2	
gi 63465 emb V00414.1 GGHI01 Gallus gallus mRNA coding for ...	30
4.2	
gi 63444 emb X13894.1 GGH2AF Chicken histone H2A.F gene	30
4.2	
Alignments	

>gi|10863872|ref|NM_000660.1| Homo sapiens transforming growth factor, beta 1

(Camurati-Engelmann disease) (TGFB1), mRNA

Length = 2745

Score = 32.2 bits (16), Expect = 1.1

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Subjct: 1 acctccctccgcggag 16

>gi|18590091|ref|XM_085882.1| Homo sapiens similar to transforming growth factor, beta 1 (H.

sapiens) (LOC147760), mRNA

Length = 697

Score = 32.2 bits (16), Expect = 1.1

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Subjct: 7 acctccctccgcggag 22

>gi|11424057|ref|XM_008912.1| Homo sapiens transforming growth factor, beta 1 (TGFB1), mRNA

Length = 2741

Score = 32.2 bits (16), Expect = 1.1

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Sbjct: 1 acctccctccgcggag 16

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)

Posted date: Apr 9, 2002 10:59 AM

Number of letters in database: 1,002,800,820

Number of sequences in database: 1,205,903

Lambda K H

1.37 0.711 1.31

Gapped

Lambda K H

1.37 0.711 1.31

Matrix: blastn matrix:1 -3

Gap Penalties: Existence: 5, Extension: 2

Number of Hits to DB: 6901

Number of Sequences: 1205903

Number of extensions: 6901

Number of successful extensions: 1479

Number of sequences better than 10.0: 16

length of query: 16

length of database: 5,297,768,116

effective HSP length: 15

effective length of query: 1
effective length of database: 5,279,679,571
effective search space: 5279679571
effective search space used: 5279679571
T: 0
A: 30
X1: 6 (11.9 bits)
X2: 15 (29.7 bits)
S1: 12 (24.3 bits)
S2: 15 (30.2 bits)

Top of Form

1: NM_000660. Homo sapiensRelated Sequences, OMIM, Protein, PubMed,
tran..[gi:10863872] Taxonomy, UniSTS, LinkOut
LOCUS NM_000660 2745 bp mRNA linear PRI 13-F
EB-2002
DEFINITION Homo sapiens transforming growth factor, beta 1 (Camurati-Engelmann
disease) (TGFB1), mRNA.
ACCESSION NM_000660
VERSION NM_000660.1 GI:10863872
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2745)

AUTHORS Derynck, R., Jarrett, J.A., Chen, E.Y., Eaton, D.H., Bell, J.R., Assoian, R.K., Roberts, A.B., Sporn, M.B. and Goeddel, D.V.

TITLE Human transforming growth factor-beta complementary DNA sequence
and expression in normal and transformed cells

JOURNAL Nature 316 (6030), 701-705 (1985)

MEDLINE 85296301

REFERENCE 2 (bases 1 to 2745)

AUTHORS Sporn, M.B., Roberts, A.B., Wakefield, L.M. and Assoian, R.K.

TITLE Transforming growth factor-beta: biological function and chemical
structure

JOURNAL Science 233 (4763), 532-534 (1986)

MEDLINE 86261803

PUBMED 3487831

REFERENCE 3 (bases 1 to 2745)

AUTHORS Chang, N.S., Mattison, J., Cao, H., Pratt, N., Zhao, Y. and Lee, C

TITLE Cloning and characterization of a novel transforming growth
factor-beta1-induced TIAF1 protein that inhibits tumor necro-
sis
factor cytotoxicity

JOURNAL Biochem. Biophys. Res. Commun. 253 (3), 743-749 (1998)

MEDLINE 99119079

PUBMED 9918798

REFERENCE 4 (bases 1 to 2745)

AUTHORS Ghadami, M., Makita, Y., Yoshida, K., Nishimura, G., Fukushima, Y
., Wakui, K., Ikegawa, S., Yamada, K., Kondo, S., Niikawa, N. and To

mita, H.

TITLE Genetic mapping of the Camurati-Engelmann disease locus to chromosome 19q13.1-q13.3

JOURNAL Am. J. Hum. Genet. 66 (1), 143-147 (2000)

MEDLINE 20100617

PUBMED 10631145

REFERENCE 5 (bases 1 to 2745)

AUTHORS Vaughn, S.P., Broussard, S., Hall, C.R., Scott, A., Blanton, S.H.

,

Milunsky, J.M. and Hecht, J.T.

TITLE Confirmation of the mapping of the Camurati-Engelmann locus to 19q13. 2 and refinement to a 3.2-cM region

JOURNAL Genomics 66 (1), 119-121 (2000)

MEDLINE 20304762

PUBMED 10843814

REFERENCE 6 (bases 1 to 2745)

AUTHORS Lim, J.M., Kim, J.A., Lee, J.H. and Joo, C.K.

TITLE Downregulated expression of integrin alpha6 by transforming growth factor-beta(1) on lens epithelial cells in vitro

JOURNAL Biochem. Biophys. Res. Commun. 284 (1), 33-41 (2001)

MEDLINE 21268957

PUBMED 11374867

COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from X02812.

1.

FEATURES Location/Qualifiers

source 1..2745
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.1"
gene 1..2745
/gene="TGFB1"
/note="TGFB; DPDI; CED"
/db_xref="LocusID:7040"
/db_xref="MIM:190180"
misc_feature 37..113
/note="pot. hairpin loops-forming region"
variation 72
/allele="-"
/allele="C"
/db_xref="dbSNP:1800999"
variation 79
/allele="-"
/allele="C"
/db_xref="dbSNP:1799753"
CDS 842..2017
/gene="TGFB1"
/note="transforming growth factor, beta 1; diaphyse
al
dysplasia 1, progressive (Camurati-Engelmann diseas
e)"
/codon_start=1
/db_xref="LocusID:7040"
/db_xref="MIM:190180"

/product="transforming growth factor, beta 1
(Camurati-Engelmann disease)"
/protein_id="NP_000651.1"
/db_xref="GI:10863873"
/translation="MPPSGLRLLPLLLPLLWLLVLTPGPPAAGLSTCKTID

MELVKRK

RIEAIRGQILSKRLASPPSQEVPPGPLPEAVLALYNSTRDRVAGESAEPEPEPEAD
YYAKEVTRVLMVETHNEIYDKFKQSTHSIYMFNTSELREAVPEPVLLSRAELRLLRR
LKLKVEQHVELYQKYSNNSWRYLSNRLLAPSDSPEWLSFDVTGVVRQWLSRGGEIEGF
RLSAHCSCDSRDNTLQVDINGFTGRRGDLATIHGMNRPFLLL MATPLERAQHLQSSR
HRRALDTNYCFSSTEKNCCVRQLYIDFRKDLGWKWIHEPKGYHANFCLGPCPYIWSLD
TQYSKVLALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPVQEQLSNMIVRSCKCS"

misc_feature 863..910

/note="pot. core sequence of signal peptide (aa -27

2 to

-257)"

variation 870

/allele="C"

/allele="T"

/db_xref="dbSNP:1982073"

variation 915

/allele="C"

/allele="G"

/db_xref="dbSNP:1800471"

misc_feature 938..1600

/note="TGFb_propeptide; Region: TGF-beta propeptide

"

misc_feature 953

/note="pot. altern. translation start site"

misc_feature 1035..1043
/note="put. glycosylation site"
misc_feature 1247..1255
/note="put. glycosylation site"
misc_feature 1370..1378
/note="put. glycosylation site"
variation 1632
/allele="C"
/allele="T"
/db_xref="dbSNP:1800472"
mat_peptide 1679..2014
/product="mature TGF-beta (aa 1-112)"
misc_feature 1715..2014
/note="TGF-beta; Region: Transforming growth factor
beta
like domain"
misc_feature 1721..2014
/note="TGFB; Region: Transforming growth factor-bet
a
(TGF-beta) family"
misc_feature 2018..2096
/note="GC-rich region"
promoter 2097..2103
/note="TATA-box-like region"
misc_feature 2517..2522
/note="put. polyadenylation signal"
polyA_site 2539
/note="polyadenylation site"
BASE COUNT 527 a 938 c 801 g 479 t

ORIGIN

1 acctccctcc gcggagc cagacagcga gggcccccggc cgggggcagg ggggacg
ccc
61 cgtccggggc accccccccg gctctgagcc gcccgcgggg ccggcctcgg cccggag
cgg
121 aggaaggagt cgccgaggag cagcctgagg ccccagagtc tgagacgagc cgccgcc
gcc
181 cccgccactg cggggaggag ggggaggagg agcgggagga gggacgagct ggtcggg
aga
241 agagaaaaaa aactttttag actttccgt tgccgctggg agccggaggc gcgggg
cct
301 cttggcgcga cgctgccccg cgaggaggca ggacttgggg accccagacc gcctccc
ttt
361 gcccggggg acgcttgctc cctccctgcc ccctacacgg cgtccctcag gcgc
cat
421 tccggaccag ccctcggag tcgcccaccc ggcctccgc aaagacttt ccccaga
cct
481 cgggcgcacc ccctgcacgc cgccctcatc cccggcctgt ctcctgagcc cccgc
atc
541 ctagaccctt tctcctccag gagacggatc tctctccgac ctgccacaga tccccta
ttc
601 aagaccaccc accttctggt accagatcgc gcccatctag gttattccg tggata
ctg
661 agacacccccc ggtccaagcc tcccctccac cactgcgc cttctccctga ggagc
cag
721 cttccctcg aggccctcct acctttgcc gggagacccc cagccctgc agggc
ggg
781 cctccccacc acaccagccc tggcgct ctcggcagtg cggggggcgc ccgc
ccc

841 catgccgcc tccgggctgc ggctgctgcc gctgctgcta ccgctgctgt ggctact
ggt
901 gctgacgcct ggcccgccgg ccgcgggact atccacctgc aagactatcg acatgga
gct
961 ggtgaagcgg aagcgcatcg aggccatccg cggccagatc ctgtccaagc tgccggct
cgc
1021 cagccccccg agccaggggg aggtgccgcc cggcccgctg cccgaggccg tgctcgc
cct
1081 gtacaacagc acccgcgacc gggtgccgg ggagagtgc aaccggagc ccgagcc
tga
1141 ggccgactac tacgccaagg aggtcacccg cgtgctaatt gtggaaaccc acaacga
aat
1201 ctatgacaag ttcaaggcaga gtacacacag catatatatg ttcttcaaca catcaga
gct
1261 ccgagaagcg gtacctgaac ccgtgttgct ctccggca gagctgcgtc tgctgag
gag
1321 gctcaagtta aaagtggagc agcacgtgga gctgtaccag aaatacagca acaattc
ctg
1381 gcgataccctc agcaaccggc tgctggcacc cagcgactcg ccagagtgg tatcttt
tga
1441 tgcaccggc gttgtcgcc agtgggtgag ccgtggaggg gaaattgagg gctttcg
cct
1501 tagcgcccac tgctcctgtg acagcaggaa taacacactg caagtggaca tcaacgg
gtt
1561 cactaccggc cgccgaggtg acctggccac cattcatggc atgaaccggc ctttcct
gct
1621 tctcatggcc accccgctgg agagggccca gcatctgcaa agctccggc accggcc
agc
1681 cctggacacc aactattgct tcagctccac ggagaagaac tgctgcgtgc ggcagct

gta

1741 cattgacttc cgcaaggacc tcggctggaa gtggatccac gagcccaagg gctacca

tgc

1801 caacttctgc ctggggccct gccctacat ttggagccctg gacacgcagt acagcaa

ggt

1861 cctggccctg tacaaccaggc ataacccggg cgcctcggcg ggcgcgtgct gcgtgcc

gca

1921 ggcgctggag ccgctgccc tcgtgtacta cgtgggcccgc aagcccaagg tggagca

gct

1981 gtccaaacatg atcgtgcgct cctgcaagt cagctgaggt cccgccccgc cccgccc

cgc

2041 cccggcaggc ccggcccccac cccgccccgc ccccgctgcc ttgccccatgg gggctgt

att

2101 taaggacacc gtggcccaag cccacctggg gccccattaa agatggagag aggactg

cgg

2161 atctctgtgt cattggcgcc ctgcctgggg tctccatccc tgacgttccc ccactcc

cac

2221 tccctctctc tccctctctg ctcctcctg cctgtctgca ctattccttt gcccggc

atc

2281 aaggcacagg ggaccagtgg ggaacactac tgttagttaga tctatttatt gagcacc

ttg

2341 ggcactgttg aagtgcctta cattaatgaa ctcattcagt caccatagca acactct

gag

2401 atggcaggga ctctgataac acccatttt aaggttgagg aaacaagccc agagagg

tta

2461 agggaggagt tcctgcccac caggaacctg cttagtgggg ggatagtgaa gaagaca

ata

2521 aaagatagta gttcaggcca ggcgggtgc tcacgcctgt aatcctagca cttttgg

gag

2581 gcagagatgg gaggatactt gaatccaggc atttgagacc agcctggta acatagt
gag

2641 acccttatctc tacaaaacac ttttaaaaaa tgtacacctg tggtcccagc tactctg
gag

2701 gctaagggtgg gaggatcact tgatcctggg aggtcaaggc tgcag

//

Bottom of Form

Revised: October 24, 2001.

Query= (16 letters)

Database: GenBank Human EST entries

4,280,058 sequences; 2,114,234,064 total letters

Score E

Sequences producing significant alignments: (bits)

Value

gi|19365764|gb|BM915385.1|BM915385 AGENCOURT_6701642 NIH_MG... 32

0.41

gi|19353768|gb|BM903897.1|BM903897 AGENCOURT_6696012 NIH_MG... 32

0.41

gi|18807810|gb|BM562052.1|BM562052 AGENCOURT_6562015 NIH_MG... 32

0.41

gi|18791603|gb|BM553137.1|BM553137 AGENCOURT_6572574 NIH_MG... 32

0.41

gi|16171065|gb|BI908151.1|BI908151 603067456F1 NIH_MGC_118 ... 32

0.41

gi|15759271|gb|BI767693.1|BI767693 603060648F1 NIH_MGC_122 ... 32

0.41

gi|15343643|gb|BI518851.1|BI518851 603061760F1 NIH_MGC_118 ... 32

0.41

gi 14309343 gb BG899094.1 BG899094	HOA21-1-G9 HOA (Human Os...	32
0.41		
gi 13662542 gb BG611171.1 BG611171	602612144F1 NIH_MGC_60 H...	32
0.41		
gi 12609210 gb BG115704.1 BG115704	602317174F1 NIH_MGC_88 H...	32
0.41		
gi 12101282 gb BF796228.1 BF796228	602258513F1 NIH_MGC_85 H...	32
0.41		
gi 11152079 gb BF238160.1 BF238160	601811886F1 NIH_MGC_48 H...	32
0.41		
gi 11100313 gb BF206727.1 BF206727	601871105F1 NIH_MGC_19 H...	32
0.41		
gi 11100272 gb BF206686.1 BF206686	601871051F1 NIH_MGC_19 H...	32
0.41		
gi 16775383 gb BM046103.1 BM046103	603625849F1 NIH_MGC_40 H...	30
1.6		
gi 19739174 gb BQ014273.1 BQ014273	UI-H-ED1-axs-h-21-0-UI. s...	28
6.4		
gi 19378603 gb BM928224.1 BM928224	AGENCOURT_6699855 NIH_MG...	28
6.4		
gi 19367808 gb BM917429.1 BM917429	AGENCOURT_6606724 NIH_MG...	28
6.4		
gi 19364214 gb BM913835.1 BM913835	AGENCOURT_6612786 NIH_MG...	28
6.4		
gi 19361343 gb BM910964.1 BM910964	AGENCOURT_6615957 NIH_MG...	28
6.4		
gi 18505954 gb BM456914.1 BM456914	AGENCOURT_6404253 NIH_MG...	28
6.4		
gi 18499709 gb BM450669.1 BM450669	AGENCOURT_6394717 NIH_MG...	28

6.4			
gi 16000196 gb BI859449.1 BI859449	603388188F1	NIH_MGC_87	H... 28
6.4			
gi 15928460 gb BI818193.1 BI818193	603032663F1	NIH_MGC_115	... 28
6.4			
gi 15431547 gb BI544235.1 BI544235	603241605F1	NIH_MGC_95	H... 28
6.4			
gi 15345229 gb BI520437.1 BI520437	603071622F1	NIH_MGC_119	... 28
6.4			
gi 14440373 gb BI033747.1 BI033747	PM3-NN0223-220201-014-h0...		28
6.4			
gi 14426676 gb BI020046.1 BI020046	CM3-MT0291-110101-622-f0...		28
6.4			
gi 14081325 gb BG770672.1 BG770672	602734012F1	NIH_MGC_49	H... 28
6.4			
gi 13546630 gb BG547965.1 BG547965	602576071F1	NIH_MGC_77	H... 28
6.4			
gi 13030375 gb BG281450.1 BG281450	602401966F1	NIH_MGC_20	H... 28
6.4			
gi 12951460 emb AL582959.1 AL582959	AL582959	LTI_NFL010_BC2...	28
6.4			
gi 12764352 gb BG254536.1 BG254536	602368464F1	NIH_MGC_91	H... 28
6.4			
gi 12378592 gb BF961317.1 BF961317	PM3-NN0223-111200-004-d0...		28
6.4			
gi 12374538 gb BF957263.1 BF957263	PM3-NN0223-241100-002-b0...		28
6.4			
gi 12323114 gb BF926150.1 BF926150	CM2-NT0193-301100-562-al...		28
6.4			

gi 12259862 gb BF869732.1 BF869732	IL3-ET0114-251000-316-A1...	28
6.4		
gi 12129894 gb BF800905.1 BF800905	PM1-CI0110-201000-003-f0...	28
6.4		
gi 12071436 gb BF744760.1 BF744760	QV2-BT0635-311000-440-c1...	28
6.4		
gi 11770407 gb BE965733.2 BE965733	601659792R1 NIH_MGC_70 H...	28
6.4		
gi 11766539 gb BE963121.2 BE963121	601656923R1 NIH_MGC_67 H...	28
6.4		
gi 10348536 gb BE890328.1 BE890328	601431783F1 NIH_MGC_72 H...	28
6.4		
gi 10142985 gb BE728993.1 BE728993	601562251F1 NIH_MGC_20 H...	28
6.4		
gi 10095527 gb BE707262.1 BE707262	PM1-HT0452-060700-008-e0...	28
6.4		
gi 9772196 gb BE543551.1 BE543551	601070523F1 NIH_MGC_12 Ho...	28
6.4		
gi 9768571 gb BE539926.1 BE539926	601060667F2 NIH_MGC_10 Ho...	28
6.4		
gi 9342607 gb BE397242.1 BE397242	601290754F1 NIH_MGC_8 Hom...	28
6.4		
gi 9332870 gb BE387505.1 BE387505	601274247F1 NIH_MGC_20 Ho...	28
6.4		
gi 8140649 gb AW950985.1 AW950985	EST363055 MAGE resequence...	28
6.4		
gi 8139665 gb AW950129.1 AW950129	EST362094 MAGE resequence...	28
6.4		
gi 6879658 gb AW375004.1 AW375004	MR0-CT0068-280999-002-f07...	28

6.4

gi|5435227|emb|AL079651.1|AL079651 DKFZp434N0629_r1 434 (sy... 28

6.4

gi|5406349|emb|AL036861.1|AL036861 DKFZp56401963_r1 564 (sy... 28

6.4

gi|2566893|gb|AA641675.1|AA641675 nr62g01.s1 NCI_CGAP_Lym3 ... 28

6.4

gi|2080087|gb|AA418268.1|AA418268 zv96d09.s1 Soares_NhHMPu_... 28

6.4

gi|2056455|gb|AA402650.1|AA402650 zu49g06.r1 Soares_ovary t... 28

6.4

gi|1516398|gb|AA040102.1|AA040102 zk46e02.r1 Soares_pregn... 28

6.4

Alignments

>gi|19365764|gb|BM915385.1|BM915385 AGENCOURT_6701642 NIH_MGC_41 Homo s
apiens cDNA clone

IMAGE:5481560 5'.

Length = 1086

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||

Sbjct: 23 acctccctccgcggag 38

>gi|19353768|gb|BM903897.1|BM903897 AGENCOURT_6696012 NIH_MGC_67 Homo s
apiens cDNA clone IMAGE:5492392

5'.

Length = 1497

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcggag 16

|||||||||||||||||

Sbjct: 445 acctccctccgcggag 430

>gi|18807810|gb|BM562052.1|BM562052 AGENCOURT_6562015 NIH_MGC_118 Homo sapiens cDNA clone

IMAGE:5745414 5'.

Length = 1175

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||||

Sbjct: 20 acctccctccgcggag 35

>gi|18791603|gb|BM553137.1|BM553137 AGENCOURT_6572574 NIH_MGC_41 Homo sapiens cDNA clone

IMAGE:5467063 5'.

Length = 1100

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||||

Sbjct: 26 acctccctccgcggag 41

>gi|16171065|gb|BI908151.1|BI908151 603067456F1 NIH_MGC_118 Homo sapien
s cDNA clone IMAGE:5216508 5'.

Length = 706

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||||

Sbjct: 25 acctccctccgcggag 40

>gi|15759271|gb|BI767693.1|BI767693 603060648F1 NIH_MGC_122 Homo sapien
s cDNA clone IMAGE:5209978 5'.

Length = 862

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 705 acctccctccgcggag 720

>gi|15343643|gb|BI518851.1|BI518851 603061760F1 NIH_MGC_118 Homo sapien
s cDNA clone IMAGE:5210943 5'.

Length = 943

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 25 acctccctccgcggag 40

>gi|14309343|gb|BG899094.1|BG899094 HOA21-1-G9 HOA (Human Osteoarthritis
Cartilage) Homo sapiens

cDNA.

Length = 364

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 83 acctccctccgcggag 98

>gi|13662542|gb|BG611171.1|BG611171 602612144F1 NIH_MGC_60 Homo sapiens

cDNA clone IMAGE:4737466 5'.

Length = 897

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Sbjct: 809 acctccctccgcggag 794

>gi|12609210|gb|BG115704.1|BG115704 602317174F1 NIH_MGC_88 Homo sapiens

cDNA clone IMAGE:4417482 5'.

Length = 838

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Sbjct: 51 acctccctccgcggag 66

>gi|12101282|gb|BF796228.1|BF796228 602258513F1 NIH_MGC_85 Homo sapiens

cDNA clone IMAGE:4341962 5'.

Length = 1081

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Sbjct: 7 acctccctccgcggag 22

>gi|11152079|gb|BF238160.1|BF238160 601811886F1 NIH_MGC_48 Homo sapiens

cDNA clone IMAGE:4054821 5'.

Length = 811

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Sbjct: 11 acctccctccgcggag 26

>gi|11100313|gb|BF206727.1|BF206727 601871105F1 NIH_MGC_19 Homo sapiens

cDNA clone IMAGE:4101600 5'.

Length = 888

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||||

Sbjct: 32 acctccctccgcggag 47

>gi|11100272|gb|BF206686.1|BF206686 601871051F1 NIH_MGC_19 Homo sapiens
cDNA clone IMAGE:4101517 5'.

Length = 917

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||||

Sbjct: 33 acctccctccgcggag 48

>gi|16775383|gb|BM046103.1|BM046103 603625849F1 NIH_MGC_40 Homo sapiens
cDNA clone IMAGE:5452309 5'.

Length = 869

Score = 30.2 bits (15), Expect = 1.6

Identities = 15/15 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcggag 16

|||||||||||||||||

Sbjct: 692 cctccctccgcggag 706

>gi|19739174|gb|BQ014273.1|BQ014273 UI-H-ED1-axs-h-21-0-UI.s1 NCI_CGAP_
ED1 Homo sapiens cDNA clone

IMAGE:5833028 3'.

Length = 772

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 495 cctccctccgcgga 482

>gi|19378603|gb|BM928224.1|BM928224 AGENCOURT_6699855 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5770072 5'.

Length = 1140

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 1009 cctccctccgcgga 1022

>gi|19367808|gb|BM917429.1|BM917429 AGENCOURT_6606724 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5483947 5'.

Length = 1073

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 916 acctccctccgcgg 929

>gi|19364214|gb|BM913835.1|BM913835 AGENCOURT_6612786 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5477539 5'.

Length = 1104

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 842 cctccctccgcgga 829

>gi|19361343|gb|BM910964.1|BM910964 AGENCOURT_6615957 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454547 5'.

Length = 1128

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 3 ctccctccgcggag 16

||||||||||||||

Sbjct: 883 ctccctccgcggag 870

>gi|18505954|gb|BM456914.1|BM456914 AGENCOURT_6404253 NIH_MGC_92 Homo sapiens cDNA clone

IMAGE:5583862 5'.

Length = 1813

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 29 cctccctccgcgga 42

>gi|18499709|gb|BM450669.1|BM450669 AGENCOURT_6394717 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5494366 5'.

Length = 1430

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 1150 acctccctccgcgg 1163

>gi|16000196|gb|BI859449.1|BI859449 603388188F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5396997 5'.

Length = 852

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 100 acctccctccgcgg 113

>gi|15928460|gb|BI818193.1|BI818193 603032663F1 NIH_MGC_115 Homo sapien

s cDNA clone IMAGE:5173838 5'.

Length = 683

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cttccctccgcgga 15

||||||||||||||

Sbjct: 96 cttccctccgcgga 109

>gi|15431547|gb|BI544235.1|BI544235 603241605F1 NIH_MGC_95 Homo sapiens

cDNA clone IMAGE:5284296 5'.

Length = 676

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 3 ctccctccgcggag 16

|||||||||||||

Sbjct: 39 ctccctccgcggag 26

>gi|15345229|gb|BI520437.1|BI520437 603071622F1 NIH_MGC_119 Homo sapien
s cDNA clone IMAGE:5163773 5'.

Length = 727

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

|||||||||||||

Sbjct: 505 acctccctccgcgg 492

>gi|14440373|gb|BI033747.1|BI033747 PM3-NN0223-220201-014-h04 NN0223 Ho
mo sapiens cDNA.

Length = 284

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

|||||||||||||

Sbjct: 97 acctccctccgcgg 84

>gi|14426676|gb|BI020046.1|BI020046 CM3-MT0291-110101-622-f04 MT0291 Ho
mo sapiens cDNA.

Length = 436

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 365 acctccctccgcgg 352

>gi|14081325|gb|BG770672.1|BG770672 602734012F1 NIH_MGC_49 Homo sapiens
cDNA clone IMAGE:4859546 5'.

Length = 949

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 63 acctccctccgcgg 76

>gi|13546630|gb|BG547965.1|BG547965 602576071F1 NIH_MGC_77 Homo sapiens
cDNA clone IMAGE:4704209 5'.

Length = 918

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 248 acctccctccgcgg 261

>gi|13030375|gb|BG281450.1|BG281450 602401966F1 NIH_MGC_20 Homo sapiens
cDNA clone IMAGE:4544201 5'.

Length = 782

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 417 acctccctccgcgg 430

>gi|12951460|emb|AL582959.1|AL582959 AL582959 LTI_NFL010_BC2 Homo sapiens
cDNA clone CS0DL008YA12 3 prime.

Length = 822

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 533 cctccctccgcgga 520

>gi|12764352|gb|BG254536.1|BG254536 602368464F1 NIH_MGC_91 Homo sapien
s cDNA clone IMAGE:4476902 5'.

Length = 1031

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 849 cctccctccgcgga 862

>gi|12378592|gb|BF961317.1|BF961317 PM3-NN0223-111200-004-d03 NN0223 H
omo sapiens cDNA.

Length = 277

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 89 acctccctccgcgg 76

>gi|12374538|gb|BF957263.1|BF957263 PM3-NN0223-241100-002-b08 NN0223 H
omo sapiens cDNA.

Length = 168

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 117 acctccctccgcgg 104

>gi|12323114|gb|BF926150.1|BF926150 CM2-NT0193-301100-562-a12 NT0193 Homo sapiens cDNA.

Length = 417

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cttccctccgcgga 15

||||||||||||||

Sbjct: 268 cttccctccgcgga 255

>gi|12259862|gb|BF869732.1|BF869732 IL3-ET0114-251000-316-A11 ET0114 Homo sapiens cDNA.

Length = 278

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 73 acctccctccgcgg 60

>gi|12129894|gb|BF800905.1|BF800905 PM1-CI0110-201000-003-f08 CI0110 Homo sapiens cDNA.

Length = 283

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 211 acctccctccgcgg 224

>gi|12071436|gb|BF744760.1|BF744760 QV2-BT0635-311000-440-c11 BT0635 Homo sapiens cDNA.

Length = 534

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 319 cctccctccgcgga 332

>gi|11770407|gb|BE965733.2|BE965733 601659792R1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3896134 3'.

Length = 1336

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 292 acctccctccgcgg 305

>gi|11766539|gb|BE963121.2|BE963121 601656923R1 NIH_MGC_67 Homo sapiens

cDNA clone IMAGE:3865924 3'.

Length = 1442

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 3 ctccctccgcggag 16

||||||||||||||

Sbjct: 403 ctccctccgcggag 390

>gi|10348536|gb|BE890328.1|BE890328 601431783F1 NIH_MGC_72 Homo sapiens

cDNA clone IMAGE:3916820 5'.

Length = 794

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 115 acctccctccgcgg 128

>gi|10142985|gb|BE728993.1|BE728993 601562251F1 NIH_MGC_20 Homo sapiens
cDNA clone IMAGE:3831924 5'.

Length = 840

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 397 acctccctccgcgg 410

>gi|10095527|gb|BE707262.1|BE707262 PM1-HT0452-060700-008-e08 HT0452 Ho
mo sapiens cDNA.

Length = 592

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 343 acctccctccgcgg 330

>gi|9772196|gb|BE543551.1|BE543551 601070523F1 NIH_MGC_12 Homo sapiens

cDNA clone IMAGE:3456940 5'.

Length = 1035

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 332 acctccctccgcgg 345

>gi|9768571|gb|BE539926.1|BE539926 601060667F2 NIH_MGC_10 Homo sapiens

cDNA clone IMAGE:3447161 5'.

Length = 902

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 411 acctccctccgcgg 424

>gi|9342607|gb|BE397242.1|BE397242 601290754F1 NIH_MGC_8 Homo sapiens c

DNA clone IMAGE:3621253 5'.

Length = 524

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 228 cctccctccgcgga 241

>gi|9332870|gb|BE387505.1|BE387505 601274247F1 NIH_MGC_20 Homo sapiens

cDNA clone IMAGE:3615538 5'.

Length = 637

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 422 acctccctccgcgg 435

>gi|8140649|gb|AW950985.1|AW950985 EST363055 MAGE resequences, MAGA Hom

o sapiens cDNA.

Length = 638

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 273 cctccctccgcgga 286

>gi|8139665|gb|AW950129.1|AW950129 EST362094 MAGE resequences, MAGA Homo sapiens cDNA.

Length = 611

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 273 cctccctccgcgga 286

Database: GenBank Human EST entries

Posted date: Mar 29, 2002 2:35 AM

Number of letters in database: 2,114,234,064

Number of sequences in database: 4,280,058

Lambda K H

1.37 0.711 1.31

Gapped

Lambda K H

1.37 0.711 1.31

Matrix: blastn matrix:1 -3

Gap Penalties: Existence: 5, Extension: 2

Number of Hits to DB: 5013

Number of Sequences: 4280058

Number of extensions: 5013

Number of successful extensions: 5013

Number of sequences better than 10.0: 61

length of query: 16

length of database: 2,114,234,064

effective HSP length: 15

effective length of query: 1

effective length of database: 2,050,033,194

effective search space: 2050033194

effective search space used: 2050033194

T: 0

A: 30

X1: 6 (11.9 bits)

X2: 15 (29.7 bits)

S1: 12 (24.3 bits)

S2: 14 (28.2 bits)

Top of Form

1: BM915385.AGENCOURT_6701642...[gi:19365764]

Taxono

my,

LinkO

ut

IDENTIFIERS

dbEST Id: 11598757

EST name: AGENCOURT_6701642

GenBank Acc: BM915385

GenBank gi: 19365764

CLONE INFO

Clone Id: IMAGE:5481560 (5')

Plate: LLCM2006 Row: d Column: 09

DNA type: cDNA

PRIMERS

PolyA Tail: Unknown

SEQUENCE

CGCCCTGGCCATTCCTCCCACCTCCCTCCGCGGAGCAGCCAGACAGCGAGGGC

CCCG

GCCGGGGGCAGGGGGGACGCCCGTCCGGGCACCCCCCGGCTCTGAGCCGCCG

CGGG

GCCGGCCTCGGCCCGAGCGGAGGAAGGAGTCGCCGAGGAGCAGCCTGAGGCCCA

GAGT

CTGAGACGAGCCGCCGCCGCCACTGCGGGAGGAGGGGAGGAGGAGCGG

GAGG

AGGGACGAGCTGGTCGGAGAAGAGGAAAAAAACTTTGAGACTTTCCGTTGCCG

CTGG

GAGCCGGAGGCGCGGGGACCTCTTGGCGCGACGCTGCCCGCGAGGAGGCAGGACT

TGGG

GACCCCAGACCGCCTCCCTTGCCGCCGGGACGCTGCTCCCTCCCTGCCCTA

CACG

GCGTCCCTCAGGCGCCCCATTCCGGACCAGCCCTCGGAGTCGCCGACCCGGCCT

CTCG

CAAAGACTTTACCATACCTCGGGCGACCCCTCTGCACGCCCTTCATCACCGG

CCTG

TCTACTGAGCCCCCGCGGATGCCTAGACCCTTCTCCTCCGGAGACGGATCCCTC

TCCG

ACCTGCCGCAAATTCCCTATTCTGGAACACCCCCGGCTCTGGGACCCCTAATCCCC

GCCT

TTCGACGCTCCTTGGCTGGGAACTGAAGAGCCCCGGGTCGTAACCTTTCTC

TCCC

CGTTTGAAAAACATCCCCCGTTAATAAACCTTGAECTATTTCGCTTGGGCCCCC

CCCT

TACGGTTTGGCGGGCACTAACAAACATCGAGTCTCAAGCGGCGGATGCCACT

CAAG

CCTGAATACTTTGCGCGTTAGGGCGGTCTTACGCGAGTAGAGTCGGGCCTTG

ACCG

GACCTATTCAATTGGTTCCCGTGACGTGTGCGGGCGTAACGAGATATTAACCTCT

CCCG

ACACATTGTCATAAAACACCACITTCGACACGCCACTCCTGTTAATAGTCGCC

CCTC

CCCGCGTGTAAAATTCCCGCGCAATGCCCTCCATTATTCCGCTCCATGAAAAAG

GGGG

TCGGCN

Quality: High quality sequence stops at base: 467

Entry Created: Mar 11 2002

Last Updated: Mar 12 2002

COMMENTS

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL

)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information c
an

be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

LIBRARY

Lib Name: NIH_MGC_41
Organism: Homo sapiens
Organ: skin
Tissue type: amelanotic melanoma, cell line
Lab host: DH10B (phage-resistant)
Vector: pOTB7
R. Site 1: XhoI
R. Site 2: EcoRI
Description: cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGA
G(G

). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using

ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT

(Life Technologies). Note: this is a NIH_MGC Library.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Institutes of Health, Mammalian Gene Collection (MGC)
Authors: NIH-MGC <http://mgc.nci.nih.gov/>
Year: 1999
Status: Unpublished

Bottom of Form

Revised: October 24, 2001.

Check on Est in Genbank:

Query= (1086 letters)

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)

1,205,903 sequences; 5,297,768,116 total letters

Score E

Sequences producing significant alignments: (bits)

Value

gi|10863872|ref|NM_000660.1| Homo sapiens transforming grow... 587

e-165

gi|18590091|ref|XM_085882.1| Homo sapiens similar to transf... 587

e-165

gi|11424057|ref|XM_008912.1| Homo sapiens transforming grow... 587

e-165

gi|7684381|gb|AC011462.4|AC011462 Homo sapiens chromosome 1... 587

e-165

gi|37097|emb|X05839.1|HSTGFBG1 Human transforming growth fa... 587

e-165

gi|37092|emb|X02812.1|HSTGFB1 Human mRNA for transforming g... 587

e-165

gi|340526|gb|J04431.1|HUMTGFBI PR Homo sapiens transforming ... 587

e-165

gi|12654682|gb|BC001180.1|BC001180 Homo sapiens, Similar to... 291

8e-76

gi 12652748 gb BC000125.1 BC000125	Homo sapiens, Similar to...	291
8e-76		
gi 18490115 gb BC022242.1	Homo sapiens, clone MGC:22008 IM...	153
4e-34		
gi 755044 gb M23703.1 PIGTGFB1A	Sus scrofa transforming gro...	129
6e-27		
gi 7650477 gb AF249327.1 AF249327	Rattus norvegicus TGF-bet...	66
8e-08		
gi 4416081 gb AF105069.1 AF105069	Rattus norvegicus transfo...	66
8e-08		
gi 2394170 gb AF015683.1 AF015683	Rattus norvegicus transfo...	66
8e-08		
gi 6755774 ref NM_011577.1	Mus musculus transforming growt...	64
3e-07		
gi 1161133 gb L42456.1 MUSTGF1G01	Mus musculus TGF-1 gene, ...	64
3e-07		
gi 3688423 emb AJ009862.1 MMU009862	Mus musculus mRNA for t...	64
3e-07		
gi 201947 gb M57902.1 MUSTGFB1	Mouse transforming growth fa...	64
3e-07		
gi 18042365 gb AC097483.3	Homo sapiens BAC clone RP11-146N...	44
0.30		
gi 17481821 ref XM_008785.3	Homo sapiens one cut domain, f...	44
0.30		
gi 12737997 ref XM_007116.2	Homo sapiens Zic family member...	44
0.30		
gi 6005961 ref NM_007129.1	Homo sapiens Zic family member ...	44
0.30		
gi 11065969 gb AF193855.1 AF193855	Homo sapiens zinc finger...	44

0.30
gi|4758847|ref|NM_004852.1| Homo sapiens one cut domain, fa... 44
0.30
gi|15787728|emb|AL355338.33|AL355338 Human DNA sequence fro... 44
0.30
gi|4028591|gb|AF104902.1|AF104902 Homo sapiens ZIC2 protein... 44
0.30
gi|1531593|gb|U50523.1|HSU50523 Human BRCA2 region, mRNA se... 44
0.30
gi|4468940|emb|Y18198.1|HSAY18198 Homo sapiens mRNA for ONE... 44
0.30
gi|19067958|gb|AY049805.1| Alopas pelagicus 5.8S ribosomal... 42
1.2
gi|18025465|gb|AY037858.1| Cercopithicine herpesvirus 15 st... 42
1.2
gi|12039248|gb|AC020659.5|AC020659 Homo sapiens chromosome ... 42
1.2
gi|19909461|gb|AC098709.3| Mus musculus clone RP23-1K14, co... 40
4.6
gi|19921137|ref|NM_135651.1| Drosophila melanogaster (CG47... 40
4.6
gi|18376846|gb|AC092198.2| Homo sapiens chromosome X clone ... 40
4.6
gi|18467841|ref|XM_078995.1| CG4751 (CG4751), mRNA 40
4.6
gi|18376869|gb|AC091898.2| Homo sapiens chromosome 5 clone ... 40
4.6
gi|18030132|gb|AC026695.5| Homo sapiens chromosome 5 clone ... 40
4.6

gi 15887302 gb AC020914.8 Homo sapiens chromosome 19 clone...	40
4.6	
gi 14578122 gb AC092241.1 AC092241 Drosophila melanogaster,...	40
4.6	
gi 15292266 gb AY051978.1 Drosophila melanogaster LD44770 ...	40
4.6	
gi 15055218 gb AC060226.39 Homo sapiens 12 BAC RP11-101P14...	40
4.6	
gi 14389338 gb AC084282.6 AC084282 Oryza sativa chromosome ...	40
4.6	
gi 13677167 gb AC015977.9 AC015977 Homo sapiens clone RP11-...	40
4.6	
gi 9910225 ref NM_020179.1 Homo sapiens FN5 protein (FN5),...	40
4.6	
gi 10440613 gb AC069145.5 AC069145 Oryza sativa chromosome ...	40
4.6	
gi 10728714 gb AE003631.2 AE003631 Drosophila melanogaster ...	40
4.6	
gi 9246422 gb AF197137.1 AF197137 Homo sapiens FN5 protein ...	40
4.6	
gi 4190938 gb AC000091.1 AC000091 Homo sapiens Chromosome 2...	40
4.6	
gi 17431932 emb AL646085.1 AL646085 Ralstonia solanacearum ...	40
4.6	
gi 15073719 emb AL591785.1 SME591785 Sinorhizobium meliloti...	40
4.6	
gi 3628578 gb AC005115.1 AC005115 Drosophila melanogaster D...	40
4.6	
gi 3150432 gb U50080.1 LSU50080 Lymnaea stagnalis serotonin...	40

4.6

gi|8052359|emb|AL356592.1|SC9H11 Streptomyces coelicolor co... 40

4.6

gi|6624640|emb|AL034344.24|HS118B18 Human DNA sequence from... 40

4.6

gi|15528721|dbj|AP003296.3| Oryza sativa (japonica cultivar... 40

4.6

gi|15289781|dbj|AP003141.2| Oryza sativa (japonica cultivar... 40

4.6

gi|6069643|dbj|AP000616.1| Oryza sativa (japonica cultivar-... 40

4.6

gi|960285|gb|L46862.1|RATLAMB2G Rattus norvegicus laminin B... 40

4.6

gi|198704|gb|J03749.1|MUSLAMB2B Mouse laminin B2 gene, exon... 40

4.6

gi|198702|gb|J02930.1|MUSLAMB2A Mouse laminin B2 chain mRNA... 40

4.6

gi|198694|gb|J03484.1|MUSLAM2B Mouse laminin B2 chain mRNA,... 40

4.6

Alignments

>gi|10863872|ref|NM_000660.1| Homo sapiens transforming growth factor,

beta 1 (Camurati-Engelmann

disease) (TGFB1), mRNA

Length = 2745

Score = 587 bits (296), Expect = e-165

Identities = 356/377 (94%), Gaps = 1/377 (0%)

Strand = Plus / Plus

Query: 246 cgagctggtcgggagaagaggnnnnnnctttgagactttccgttgccgtggagcc
305

||||||||||||||||||||| |||||||||||||||||||||||

Sbjct: 225 cgagctggtcgggagaagaggaaaaactttgagactttccgttgccgtggagcc
284

Query: 306 ggaggcgcgggacctcttggcgacgctgccccgaggaggcaggacttgggaccc
365

|||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct: 285 ggaggcgcgggacctcttggcgacgctgccccgaggaggcaggacttgggaccc
344

Query: 366 cagaccgcctcccttgcgcgggacgcttgcctccctgcacacggcgtc
425

|||||||||||||||||||||||||||||||||||||||||||||

Sbjct: 345 cagaccgcctcccttgcgcgggacgcttgcctccctgcacacggcgtc
404

Query: 426 cctcaggcgccccattccggaccagccctcgaggactcgccgacccggctctcgcaaag
485

|||||||||||||||||||||||||||||||||||||

Sbjct: 405 cctcaggcgccccattccggaccagccctcgaggactcgccgacccggctccgcaaag
464

Query: 486 actttcaccatacctcgccgcacccttgcacgcggcattcatcaccggctgtctac
545

||||||| ||| ||||||||||| ||| ||| ||| ||| ||| ||| |||

Sbjct: 465 actttccccagacacctcgccgcacccctgcacgcggcattcatccggctgtctcc
524

Query: 546 tgagccccgcggatgcctagacccttctcctccggagacggatccctctccgacctg
605

||||||||||| ||||||||||||||| ||||||||||| |||||||||

Sbjct: 525 tgagccccgcgcat-cctagacccttctcctccaggagacggatctctccgacctg
583

Query: 606 ccgcaaattccctattc 622

|| || || |||||

Sbjct: 584 ccacagatcccatttc 600

【0083】

Example 6: Statistical analysis of 5' end sequence tags

5' end sequence tags obtained from the same plurality of mRNAs in a sample or nucleic acid fragments within the same cDNA library can be analyzed by a standard software solution like NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) to identify non-redundant sequence tags as described in Example 5. All such non-redundant sequence tags can then be individually counted and further analyzed for the contribution of each non-redundant tag to the total number of all tags obtained from the same sample. The contribution of an individual tag to the total number of all tags should allow for a quantification of the transcripts in a plurality of mRNAs in the sample or a cDNA library. The results obtained in such a way on individual samples can be further compared with similar data obtained from other samples to compare their expression patterns.

【0084】

Example 7: Mapping of 5' end sequence tags to genomic sequence information

5' end specific sequence tags obtained as described in this Example can be used to identify transcribed regions within genomes for which partial

l or entire sequences were obtained. Such a search can be performed using standard software solutions like NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) to align the 5' end specific sequence tags to genomic sequences. In the case of large genomes like those from human, rat or mouse it may be necessary to extend the initial sequence information obtained from concatamers for example by the approach described in Example #. The use of extended sequences allows for a more precise identification of actively transcribed regions in the genome.

【0085】

Example 8: Identification of transcriptional start sites

5' end specific sequence tags, which could be mapped to genomic sequences, allow for the identification of regulatory sequences. In a gene the DNA upstream of the 5' end of transcribed regions usually encompasses most of the regulatory elements, which are used in the control of gene expression. These regulatory sequences can be further analyzed for their functionality by searches in databases, which hold information on binding sites for transcription factors. Publicly available databases on transcription factor binding sites and for promoter analysis include:

Transcription Regulatory Region Database (TRRD) (<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trrd4/>)

TRANSFAC (<http://transfac.gbf.de/TRANSFAC/>)

TFSEARCH (<http://www.cbrc.jp/research/db/TFSEARCH.html>)

PromoterInspector provided by Genomatix Software (<http://www.genomatix.de/>)

【0086】

Example 9: Cloning of full-length cDNAs using information derived from 5' end sequence tags

Sequence information derived from the concatamers can be used to synth

esis specific primers for the cloning of full-length cDNAs. In such an approach, the sequence derived from a given 5' end specific tag can be used to design a forward primer while the choice of the reverse primer would be dependent on the template DNA used in the amplification reaction. Amplification by the polymerase chain reaction (PCR) can be performed using a template derived from a plurality of RNA obtained from a biological sample and an oligo-dT primer. In the first step the oligo-dT primer and a reverse transcriptase are used to synthesize a cDNA pool. In the second step a forward primer derived from a 5' end specific tag and an oligo-dT primer are used to amplify a full-length cDNA from the cDNA pool. Similarly, a specific full-length cDNA can be amplified from an existing cDNA library using a forward primer derived from a 5' end tag and a vector nested reversed primer.

【0087】

Example 10: Alternative approaches for the cloning of 5'-end tags from cDNA libraries.

A plurality of cDNAs can be amplified from an existing cDNA library having a recognition site for a class IIIs endonuclease at the 5' end of the inserts. The PCR products derived from such a library would be further treated as described in the examples herein.

【0088】

Example 11: Cloning of 5' ends by replacement of the Cap structure by an oligonucleotide having a class IIIs recognition site

A cDNA/RNA hybrid encompassing the 5' end of an initial transcript can be obtained as described in Example 1. The Cap structure in such cDNA/RNA hybrids is then enzymatically removed by a hydrolyzing enzyme such as the T4 polynucleotide kinase or the tobacco acid pyrophosphatase. A single or double stranded oligonucleotide having a class IIIs recognition site is then ligated by T4 RNA ligase to the RNA at the phosphate present

at the 5' end of the de-capped mRNA. The ligated oligonucleotide will function as a primer for the second strand synthesis following the procedure given in Example 1. By the use of a modified oligonucleotide in the ligation step the double stranded cDNA can be attached to a support and used for the cloning of concatamers as described herein.

【0089】

Example 12: Amplification step for a sample

In cases where the amount of a sample is limiting to the invention, the sample material can be amplified by the following approach. In a first step a plurality of mRNAs is treated as described in Example 11 to replace the cap structure by an appropriate oligonucleotide having a class I Is recognition site. In a second step the aforementioned template is amplified by a PCR step using a primer complementary to the linker and a poly-A primer. The PCR product can be used for the invention as described in the Examples 1.

【0090】

Example 13: Utilization of extended 5'-end sequences

Initial 5' end sequences obtained for concatamers can be used to synthesis sequencing primers to obtain extended sequence information on the 5' end of a transcribed region.

【0091】

Example 14: Gene inactivation

Sequence information obtained from 5' end specific sequence tags can be used for the design of anti-sense probes, which could be applied in knockdown studies.

【0092】

【発明の効果】

By the present invention, novel means by which not only the information on the nucleotide sequences of mRNAs contained in a sample may be obta

ined, but also novel genes could be cloned. By the method of the present invention, information on the nucleotide sequences of the 5' end regions of a plurality of nucleic acids said mRNAs and cDNAs in the sample could effectively be obtained. Since the information on the nucleotide sequences of the 5' end regions is obtained, unknown genes can be cloned after the identification of a novel transcript. Further, it may be possible to attain mapping of transcription start sites, mapping of promoter usage pattern, analysis of SNPs in promoters, creating gene networking by combining the expression analysis, alternative promoter usage and the other data in this disclosure, and selective recovery of promoter regions in fragmented genomic DNA.

【0093】

In particular, the invention has a great impact on identification, cloning and further analysis of promoter regions. After sequencing concatamer libraries holding information on a plurality of 5' ends, a statistical analysis on the distribution on the transcriptional start sites will be possible. Changes between different physiological conditions switch the mRNA transcription machinery into new "status". Such a "transcriptional status" can be measured by computing (1) the presence of the transcription starting points, (2) the digital expression of the various transcriptional factors by counting their expression by counting the tags, and correlating the presence of starting point, the transcription factors. More information will be obtained on the gene networking by comparing the perturbation of gene expression between two different conditions. Such comparisons of transcriptional conditions between various disease and normal tissues could allow for the design of new and very comprehensive diagnostic tools. Thus the invention will be of high commercial value in gene discovery and gene analysis, and it is envisioned that the invention will be of use in the development of novel diagnostic and therapeutic produ

cts.

【0094】

[REFERENCES]

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW, Serial analysis of gene expression, Science 1995 Oct 20;270(5235):484-7

US patent 5,866,330 (SAGE)

US patent 5,695,937 (SAGE)

Piero CARNINCI et al., METHODS IN ENZYMOLOGY, VOL. 303, pp. 19-44, 1999

Lee S, Clark T, Chen J, Zhou G, Scott LR, Rowley JD, Wang SM, Correct identification of genes from serial analysis of gene expression tag sequences, Genomics 2002 Apr;79(4):598-602

Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE, Using the transcriptome to annotate the genome, Nat Biotechnol 2002 May;20(5):508-12

Maruyama K and Sugano S, Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene. 1994, Vol. 138:171-4

Edery I, Chu LL, Sonenberg N, Pelletier J, An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture), Mol Cell Biol 1995 Jun;15(6):3363-71

US patent 6,022,715 (GenSet)

Shibata Y, Carninci P, Watahiki A, Shiraki T, Konno H, Muramatsu M, Hayashizaki Y, Cloning full-length, cap-trapper-selected cDNAs by using the single-strand linker ligation method, Biotechniques 2001 Jun;30(6):1250-4

Sambrook J and Russel DW, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, 2001

Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watahiki A, Shibata K, Konno H, Muramatsu M, Hayashizaki Y, Balanced-size and 1

ong-size cloning of full-length, cap-trapped cDNAs into vectors of the n ovel lambda-FLC family allows enhanced gene discovery rate and functiona l analysis, *Genomics*. 2001 Sep;77(1-2):79-90.

Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignati eva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, Podkolodny NL, Kolchano v NA, Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL , *Nucleic Acids Res* 1998 Jan 1;26(1):362-7

Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*. 1994 Jan 28;138(1-2):171-4.

Jordan B., *DNA Microarrays: Gene Expression Applications*, Springer-Verla g, Berlin Heidelberg New York, 2001

Schena A, *DNA Microarrays, A Practical Approach*, Oxford University Press , Oxford 1999

US patent 5,962,272 (Clontech)

Carninci P, Shiraki T, Mizuno Y, Muramatsu M, Hayashizaki Y, Extra-long first-strand cDNA synthesis, *Biotechniques* 2002 May;32(5):984-5

US patents 6,352,828; 6,306,597; 6,280,935; 6,265,163; 5,695,934 (Lynx)

4 Brief Description of Drawings

【図1】

An example for the preparation of a plurality of 1st strand cDNAs is pr esented, where the starting material can be RNA derived from a biologica l sample or a cDNA library.

【図2】

An example for the cloning of 5' -end specific tags into concatemers is presented. The example is including but not limited to the use of the r estriction enzymes Gsu I, Bgl II and Eco RI.

【図3】

An example for a 1st linker to be used for the cloning of 5' -end speci

fic tags is presented. The example is including but not limited to the use of the restriction enzymes Bgl II, Gsu I, and Mme I.

【図4】

An example for a 2nd linker to be used for the cloning of 5' -end specific di-tags is presented. The example is including but not limited to the use of the restriction enzymes Bgl II, Gsu I, Mme I and Eco RI.

【図5】

An example for the structure of a di-tag is presented. The example is including but not limited to the use of the restriction enzymes Bgl II, Gsu I, Mme I and Eco RI.

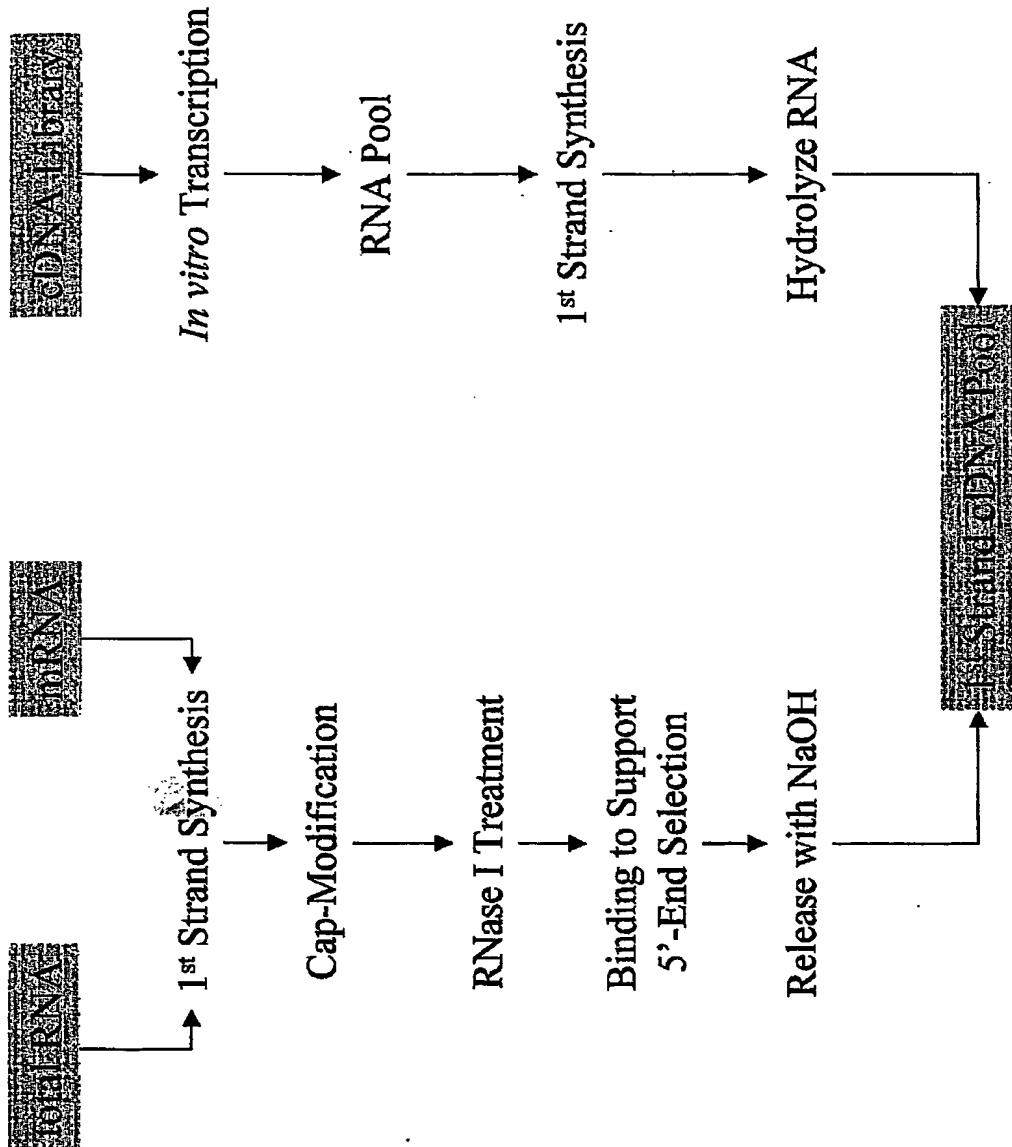
【図6】

An example for the use of a 5' -end specific linker is presented, in which the linker is used for the enrichment of individual nucleic acids and their sequencing.

【書類名】

外国語図面

【図1】

Fig.1: 1st Strand Synthesis

【図2】

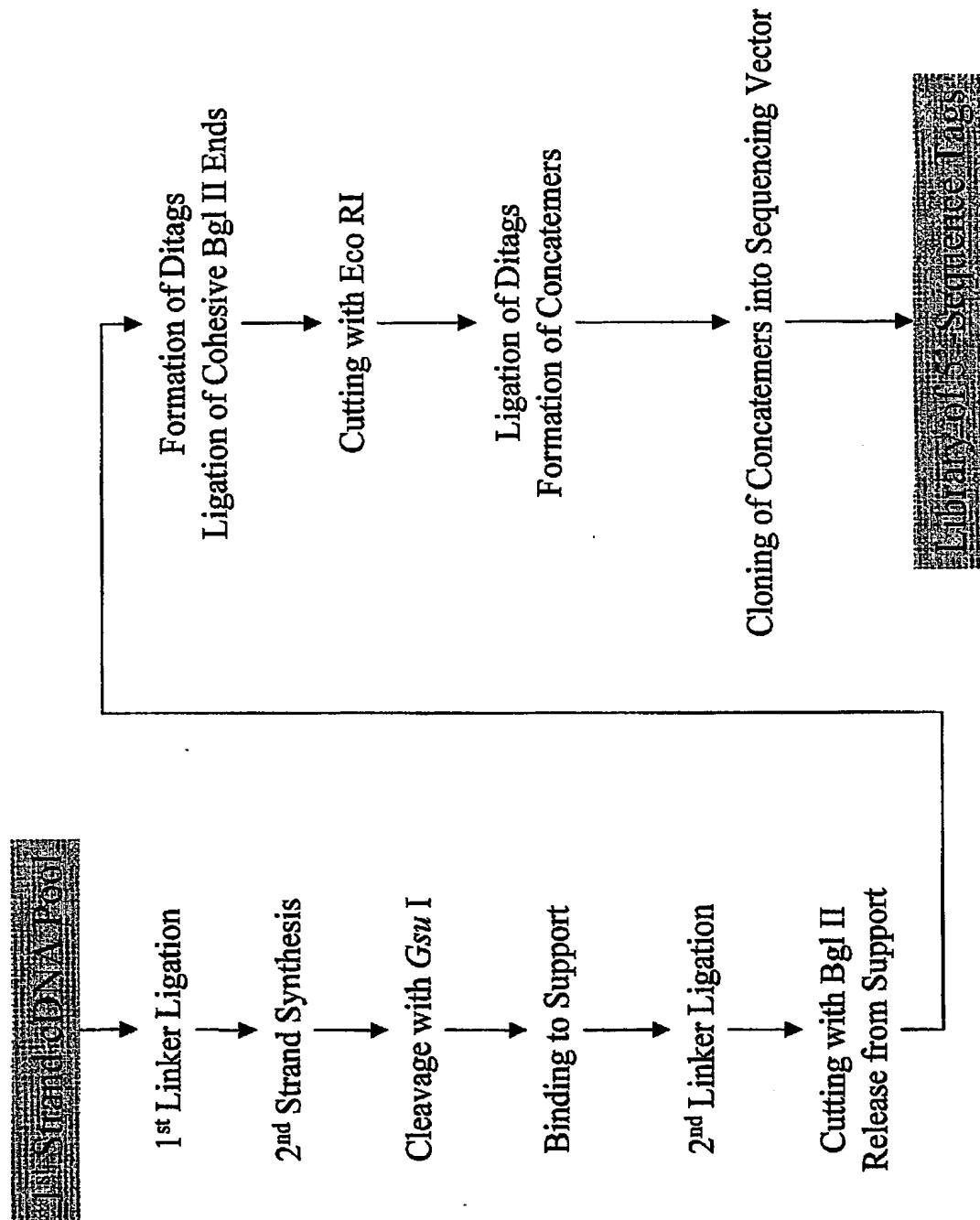
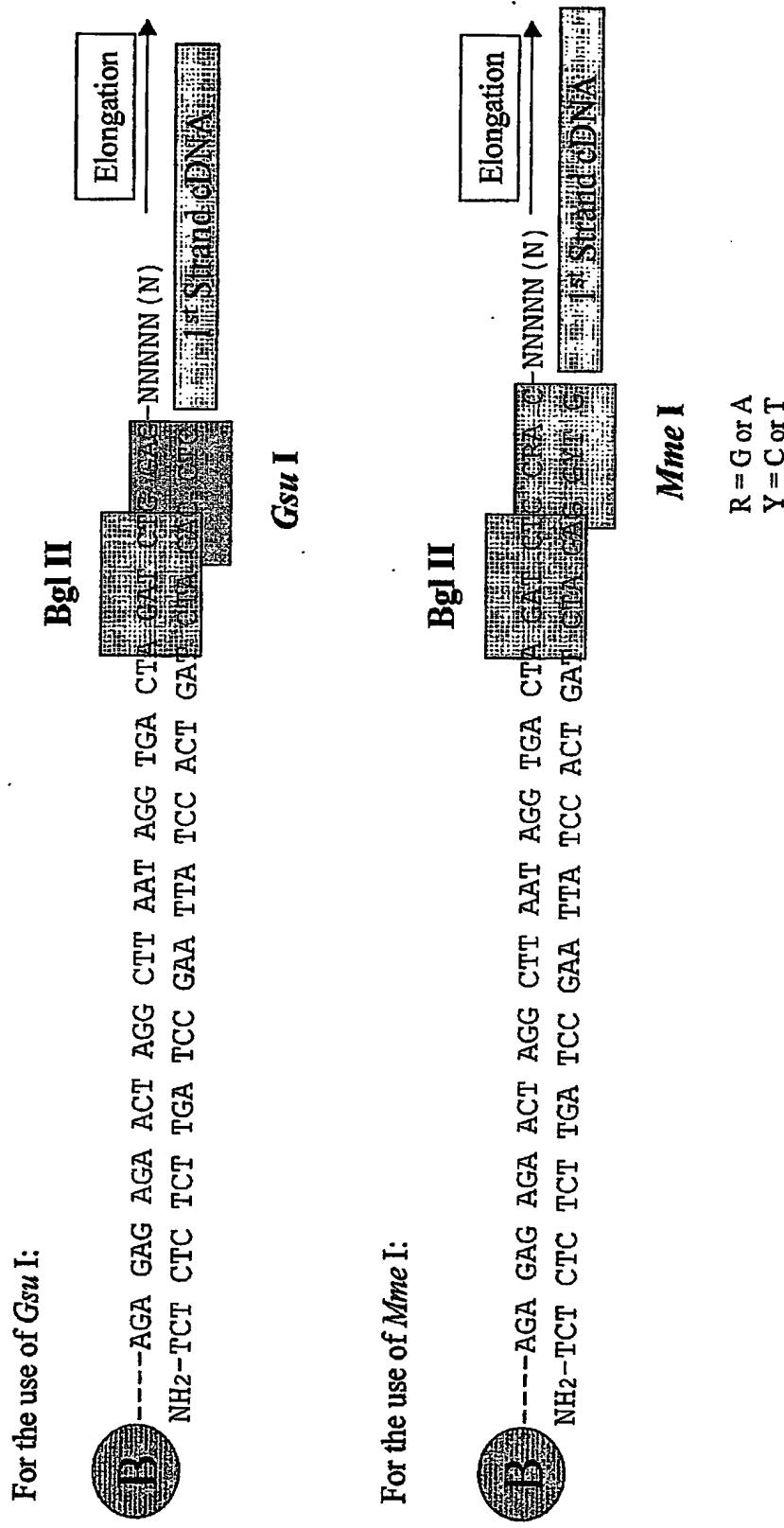


Fig.2: Cloning of Concatemers

【図3】

Fig.3: 1st Linker Ligation

【図4】

For the use of *Gsu* I:

For the use of *Mme I.*:

$$R = G \text{ or } A$$

$$Y = C \text{ or } T$$

Fig.4: 2nd Linker Ligation

【図5】

For the use of *Gsu* I:

Diagram illustrating a DNA sequence with restriction sites for Eco RI, Bgl II, and Eco RI. The sequence is: GAGGCTCCGGAGGCTCC. Eco RI sites are at positions 1 and 11. Bgl II sites are at positions 5 and 9. An arrow points to the Bgl II site at position 9.

For the use of Mine I:

$$R = G \text{ or } A$$

Fig.5: Structure of a Di-tag

【図 6】

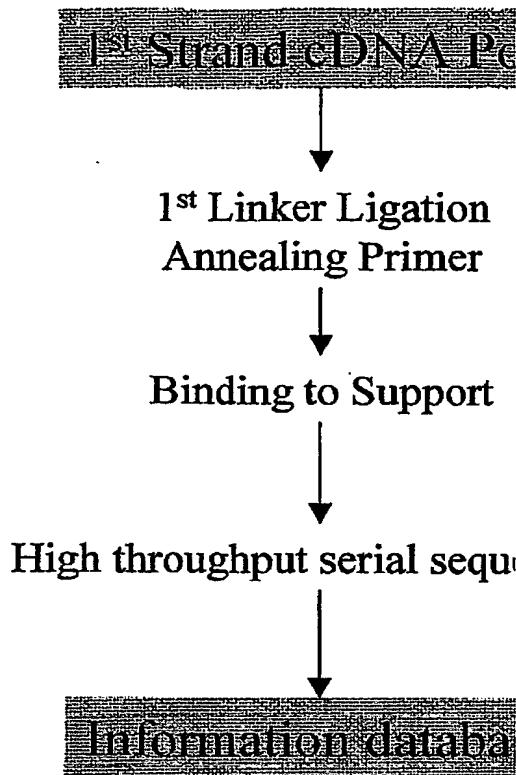


Fig.6: Serial Sequencing

【書類名】 外國語要約書

1 Abstract

A method is disclosed to obtain the 5' ends of transcribed regions from a plurality of nucleic acid fragments obtained from biological materials or synthetic pools. DNA fragments encoding the 5' ends are enriched for their individual analysis or for the analysis of concatamers thereof. The sequence information derived from 5' ends can be used for characterization and cloning of the transcriptome.

2 Representative Drawing None

認定・付加情報

特許出願の番号	特願2002-235294
受付番号	50201202416
書類名	特許願
担当官	田丸 三喜男 9079
作成日	平成15年 2月24日

<認定情報・付加情報>

【提出日】	平成14年 8月12日
【特許出願人】	
【識別番号】	000006792
【住所又は居所】	埼玉県和光市広沢2番1号
【氏名又は名称】	理化学研究所
【特許出願人】	
【識別番号】	501293666
【住所又は居所】	東京都港区三田1丁目3番35号
【氏名又は名称】	株式会社ダナフォーム
【代理人】	申請人
【識別番号】	100088546
【住所又は居所】	東京都千代田区飯田橋4丁目5番12号 岩田ビル6階 谷川国際特許事務所
【氏名又は名称】	谷川 英次郎

次頁無

【書類名】 翻訳文提出書

【あて先】 特許庁長官殿

【出願の表示】

【出願番号】 特願2002-235294

【特許出願人】

【識別番号】 000006792

【氏名又は名称】 理化学研究所

【特許出願人】

【識別番号】 502049114

【氏名又は名称】 株式会社ダナフォーム

【代理人】

【識別番号】 100088546

【弁理士】

【氏名又は名称】 谷川 英次郎

【電話番号】 03(3238)9182

【確認事項】 本書に添付した翻訳文は、外国語書面出願の願書に添付して提出した外国語明細書、外国語図面及び外国語要約書に記載した事項を過不足なく適正な日本語に翻訳したものである。

【提出物件の目録】

【物件名】 外国語明細書の翻訳文 1

【物件名】 外国語図面の翻訳文 1

【物件名】 外国語要約書の翻訳文 1

【書類名】 明細書

【発明の名称】 転写された領域の5'末端をクローニング及び分析のため
に使用する方法

【特許請求の範囲】

【請求項1】 mRNAの転写領域の5'末端に対応する核酸タグの製造方法。

【請求項2】 このような5'末端タグの連結体が製造される請求項1記載
の方法。

【請求項3】 前記mRNAの転写された領域から誘導されるこのような5
'末端特異的配列タグをシーケンシングにより分析する方法。

【請求項4】 mRNAの5'末端領域に相補的な領域を含む複数のcDNA
である、生物学的試料から誘導されたRNA若しくはmRNA、又はcDNA
から誘導された、インビトロで合成されたRNA、又は試料中のタグライブ
ラリーを鋳型として用いて形成されたしDNAを選択的に回収する第1工程と、
回収されたcDNA中の、前記mRNAの5'末端領域に相補的な領域を少なくとも
も含むcDNA領域を含む断片を選択的に回収する第2工程と、回収された断片
を互いに連結して連結体を形成する第3工程とを含む、試料中の転写に関係する
複数のmRNAの5'末端領域の塩基配列情報を含む少なくとも2個又はそれ以
上の核酸断片の連結体の作製方法。

【請求項5】 第1工程は、mRNAのキャップ構造をオリゴヌクレオチド
で置換することであり、第2工程は、cDNAの全長を形成することであり、第
3工程は、5'末端タグの開裂と連結体の形成をを含む請求項4記載の方法。

【請求項6】 前記第1工程は、mRNAを鋳型にして第一鎖cDNAを合
成する工程と、mRNAのキャップ構造に選択結合性物質を結合させる工程と、
一本鎖RNAを切断する工程と、該選択結合性物質と選択的に結合する、支持体
上に固定化された対応選択結合性物質と前記選択結合性物質とを結合させる工程
と、該支持体を回収する工程とを含む請求項4記載の方法。

【請求項7】 全長cDNAを単離する第1工程は、RNaseによる消化
工程と、それに続き、固定化されたキャップ結合性物質で処理することと、それ

に続き、このような全長cDNAを溶出することを含む請求項4記載の方法。

【請求項8】 転写物の5'末端に対応する核酸を、認識配列の外側でこのような核酸を開裂することができる物質によって認識されるように、5'末端に付加する方法。

【請求項9】 前記選択結合性物質がビオチンであり、前記対応選択結合性物質がアビジン、ストレプトアビジン又はビオチンと特異的に結合するアビジン若しくはストレプトアビジン誘導体である請求項4記載の方法。

【請求項10】 前記選択結合性物質がジゴキシゲニンであり、前記対応選択結合性物質が抗ジゴキシゲニン抗体である請求項4記載の方法。

【請求項11】 選択結合性物質は、支持体上に固定化された対応選択結合性物質に結合されており、このような支持体が磁性ビーズ、アガロースビーズ又はラテックスビーズである請求項4または9記載の方法。

【請求項12】 前記第2工程は、第1工程で回収されたcDNAの、前記mRNAの5'末端に対応する側の末端に、認識配列の外側でDNAを開裂する物質の制限酵素部位と、3'末端領域にランダムオリゴマー部分とを少なくとも含むリンカーを結合させる工程と、該リンカー又は該リンカーに部分的又は完全に対応する他のオリゴヌクレオチドをプライマーとし、前記cDNAを鋳型として第二鎖cDNAを合成する工程と、得られたリンカー結合二本鎖cDNAを前記制限酵素で処理する工程と、該制限酵素により切断されて生じる断片であって、前記リンカー部分及びcDNAの5'末端部分を含む断片を選択的に回収する工程とを含む請求項4及び6ないし11のいずれか1項に記載の方法。

【請求項13】 前記リンカーには、選択結合性物質が結合されており、リンカー部分を含む断片を選択的に回収する前記工程は、該選択結合性物質と選択的に結合する、支持体上に固定化された対応選択結合性物質と前記選択結合性物質とを結合させる工程と、該支持体を回収する工程とを含む請求項4ないし12のいずれか1項に記載の方法。

【請求項14】 前記選択結合性物質がビオチンであり、前記対応選択結合性物質がアビジン、ストレプトアビジン又はビオチンと特異的に結合するアビジン若しくはストレプトアビジン誘導体である請求項4ないし13のいずれか1項

に記載の方法。

【請求項15】 前記選択結合性物質がジゴキシゲニンであり、前記対応選択結合性物質が抗ジゴキシゲニン抗体である請求項4ないし13のいずれか1項に記載の方法。

【請求項16】 前記制限酵素は、核酸を認識し、認識部位とは異なる部位を開裂する酵素活性を有する物質である請求項4ないし15のいずれか1項に記載の方法。

【請求項17】 前記制限酵素は、Gsu I, Mme I, Bpm I又はBsg IのようなクラスII制限酵素である請求項4ないし16のいずれか1項に記載の方法。

【請求項18】 請求項1ないし18のいずれか1項に記載の方法により得られた核酸断片を用いて連結体にクローニングする工程をさらに含む方法。

【請求項19】 請求項1ないし18のいずれか1項に記載の方法により作製された連結体をシーケンスすることにより、複数のmRNAの5'末端の塩基配列を決定する方法。

【請求項20】 請求項1ないし18のいずれか1項に記載の方法において、前記第1工程に代えて、予め得られている完全長cDNAを出発材料として用いる方法。

【請求項21】 核酸を認識することができ、その認識配列の外側で核酸を開裂することができる物質により開裂可能な配列をその5'末端に有するRNA分子の混合物を既存のcDNAライブラリーから調製する、mRNAの5'末端に対応する5'末端核酸タグの製造方法。

【請求項22】 核酸を認識することができ、その認識配列の外側で核酸を開裂することができる物質により開裂可能な配列をその5'末端近傍に有する核酸タグ分子の混合物を既存のcDNAライブラリーから調製する、核酸タグ分子の製造に用いられる、mRNAの5'末端に対応する5'末端核酸タグの製造方法。

【請求項23】 請求項1ないし22のいずれか1項に記載の方法により作製された連結体。

【請求項24】 請求項23の連結体を含むベクター。

【請求項25】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から誘導される配列。

【請求項26】 ある細胞の転写状態、したがって転写ネットワーキングを決定することを可能にする、請求項1ないし22のいずれか1項に記載の方法に基づく方法。

【請求項27】 請求項1ないし22のいずれか1項に記載の方法と同じ方法であって、試料中の複数のmRNA又はcDNAについて発現データを得る方法。

【請求項28】 請求項1ないし22のいずれか1項に記載の方法と同じ方法であって、試料中の複数のmRNA又はcDNAについて発現データを定量する方法。

【請求項29】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、連結体から得られた配列情報を保持するデータベースの構築に用いる請求項28記載の方法。

【請求項30】 請求項1ないし22のいずれか1項に記載の方法と同じ方法であって、ゲノム配列中のオープンリーディングフレームを同定する方法。

【請求項31】 請求項1ないし22のいずれか1項に記載の方法と同じ方法であって、ゲノム配列中の転写開始部位及び該転写開始部位の上流の調節配列を同定する方法。

【請求項32】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、複数の核酸からの全長又は部分cDNAをクローニングすることに用いる方法。

【請求項33】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、ゲノム中の調節領域のプロモーター活性を分析することに用いる方法。

【請求項34】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、遺伝子を不活性化することに用いる方法。

【請求項35】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、前記リンカーの塩基配列を合成することに用いる方法。

【請求項36】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、前記プライマーの塩基配列を合成することに用いる方法。

【請求項37】 請求項1ないし22のいずれか1項に記載の方法により調製された連結体から得られる配列情報を、転写物の5'末端から誘導される、伸長された塩基配列を得ることに用いる方法。

【請求項38】 一本鎖cDNAを、前記リンカーの二本鎖合成オリゴヌクレオチドに連結し、該リンカーは、前記タグの塩基配列を包含する一本鎖オーバーハングを有し、該タグは、請求項1ないし13のいずれか1項に記載の方法により調製された連結体から得られたものであり、前記リンカーは選択結合性物質に結合され、選択結合性物質は、支持体上の対応選択結合性物質に連結され、このようなリンカーが前記RNA転写物である前記第一鎖cDNAの特異的塩基配列を豊富化するために用いられる、請求項1ないし8のいずれか1項に記載の方法。

【請求項39】 一本鎖cDNAを、前記プライマーの二本鎖合成オリゴヌクレオチドに連結し、選択結合性物質が前記リンカーに結合され、選択結合性物質は、支持体上の対応選択結合性物質に連結され、このようなDNA鑄型が前記RNAの最初の転写物の5'領域の塩基配列を得るために用いられる、請求項1ないし8のいずれか1項に記載の方法。

【請求項40】 診断ツールの開発に用いられる、請求項1ないし39のいずれか1項に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、試料中の複数のmRNAの5'末端領域の塩基配列情報を含む複数の核酸断片の選択的採取方法に関する。本発明の方法は、新規遺伝子の発見及び

遺伝子調節の研究に有効である。

【0002】

【従来の技術】

ゲノム情報を利用するために、ゲノムの一部がmRNAに転写される。ゲノムを理解し、それを調節プロセスに用いるためには、個々のmRNA種の情報が必要である。RNA種の情報には、その部分長又は全長の塩基配列及び特定の生物学的文脈におけるその相対量又は絶対量が含まれるべきである。

【0003】

従来、細胞や組織等の試料中に含まれるmRNAの塩基配列の解析は、mRNAを鋳型にした逆転写を利用してcDNAライブラリーを作製し、該cDNAライブラリー中の挿入cDNA断片をそれぞれ調べることにより行われていた。試料中には多数の異なるmRNAが含まれるので、この方法では、効率が悪かった。このため、複雑な試料中のmRNAの発現パターンを監視し、短い配列要素であるタグにより遺伝子を同定する他の技術が発明された。

【0004】

高効率発現プロファイリングは、一般的にいわゆるDNAマイクロアレイ (Jordan B., DNA Microarrays: Gene Expression Applications, Springer-Verlag, Berlin Heidelberg New York, 2001: Schena A, DNA Microarrays, A Practical Approach, Oxford University Press, Oxford 1999) を用いて行われている。これらの実験のために、個々の遺伝子又は転写物を表す特異プローブが支持体上に置かれ、複数の試料と同時にハイブリダイズされる。支持体上のプローブが、試料中に存在する分子と反応すると、陽性シグナルが得られるであろう。これらの実験は、多数の遺伝子又は転写物の並行的な分析を可能にする。しかしながら、このアプローチには、初めに他の実験手段により同定された遺伝子又は転写物についてのみ研究することができるという制限がある。このような手段は、cDNAライブラリー、部分配列タグ及び/又はコンピューター予測から得られた結果を包含し得る。DNAマイクロアレイ実験の

制限の故に、複数のmRNA試料から得られた部分配列であるタグに基づく方法が、遺伝子の発見及び発現プロファイリングのために用いられている。

【0005】

mRNAの塩基配列情報を効率的に得る方法として、いわゆるSAGE (Serial Analysis of Gene Expression) 法が知られている (Velculescu V. E. et al., Science 270, 484-487 (1995))。この方法は、複数のmRNAの3'末端領域の塩基配列情報を含む複数の短い (10 bp程度) DNA断片を連結して連結体DNAを構成し、この連結体DNAの塩基配列を決定するものであり、この方法によれば、多数のmRNAの3'末端領域の塩基配列情報を効率的に知ることができる。SAGE法によれば、3'末端近傍の短い領域の塩基配列しか不明であるが、mRNAが既知の場合、10 bp程度の短い領域の塩基配列がわかつただけでも統計的にmRNAをほぼ確定的に特定することが可能である。従って、この方法は、特定の細胞や組織が発現している遺伝子の解析にとって重要な方法として現在、広く用いられている。

【0006】

【発明が解決しようとする課題】

SAGE法では、mRNAの3'末端領域の塩基配列を知ることができると、3'末端領域の塩基配列情報だけでは、新規遺伝子のクローニングを行うことは困難である。SAGEは用いられているにもかかわらず、SAGEはcDNAの5'末端近傍のcDNAクローンを得る方法を教示しない。実際、4 bpのクラスIIs制限酵素が用いられている。4 bpカッターは、通常、平均して、200~300のヌクレオチドを開裂し、これはmRNA転写物の平均的なサイズの平均して1/10である。このようにSAGE原理は、3'末端については広く集めることができるけれども、ほとんどの転写物の5'末端については情報を集めることができないことを強く示唆している。さらに、10 bpのタグは、特定の遺伝子同定及びゲノムの全体又は部分の配列をマッピングするためにはしばしば不十分である。したがって、10 bpのタグは、mRNAの一部を含む「sage-tag」の同定にのみ用いられている。哺乳動物のmRNAは、哺乳動物

のゲノムの転写された部分の3～5%を含むに過ぎず、特異的「s a g e - t a g」は、分析に用いられるクラスI I a制限酵素の近傍に位置する、上記3～5%の一部を含む。4 b p制限酵素は、約44 b p (256 b p)毎にランダムに配列を切断するので、「s a g e - t a g」は、ゲノムの発現された3～5%の部分の約1/256を表すのみである（計算=0.02%未満）。したがって、S A G E技術は、ゲノムの解析にS A G E-タグを用いることはできず、非常に限定されたその一部の分析にのみ用いることができる教示している。

【0007】

従つて、本願発明の目的は、試料中に含まれるm R N Aの塩基配列情報を取得できるのみならず、新規遺伝子のクローニングを行うことも可能にする新たな手段を提供することである。

【0008】

これは、D N A転写開始部位についての統計を含むことができる。多数の5'配列の情報を得るために、本発明で提供される連結体を用いることによって、効率的に転写開始部位及び関連するプロモーター配列をマッピングすることが可能である。このように、S A G Eが無関係な3'末端を用いるためにプロモーターの解析を全く行うことができないのに対し、本発明は、新しい手段を提供する。同時に、全長c D N Aクローン及びそれらの誘導配列を集めることのできる技術が存在した。しかしながら、これらは、全長c D N Aクローンを集めることに商店が当たられており、5'末端をカバーする断片を集めるものではない。したがって、全長c D N Aをクローニングするというアプローチは、転写開始部位及び関連するプロモーター領域の同定及び分析のための高効率な方法としては適していない。本発明は、ここに、対照的な技術を組み合わせ、効率的なアプローチによって、プロモーターのマッピング及び分析に有用な5'末端を得る方法を提供する。複雑な調節ネットワークを研究し、解析するために、本発明を、新規な遺伝子を同定しクローニングする能力と共に用いることにより、生物系並びに発達、ホメオスタシス及び疾病中の状態を監視する、本発明の広範囲な用途が開かれる。

【0009】

【課題を解決するための手段】

本願発明者等は、銳意研究の結果、mRNAの5'末端領域の塩基配列情報を含む複数の核酸断片を選択的に回収することによりmRNAの塩基配列情報を取得できるのみならず、新規遺伝子のクローニングを行うことも可能であることに想到し、かつ、これを達成する具体的な方法に想到して本発明を完成した。

【0010】

すなわち、本発明は、試料中のmRNAを鑄型として形成される複数のcDNAであって、該mRNAの5'末端領域に相補的な領域を含むcDNAを選択的に回収する第1工程と、回収されたcDNA中の、前記mRNAの5'末端領域に相補的な領域と相補的な領域を少なくとも含むcDNA領域を含む断片を選択的に回収する第2工程と、回収された断片を互いに連結して連結体を形成する第3工程とを含む、試料中の複数のmRNAの5'末端領域の塩基配列情報を含む複数の核酸断片の連結体の作製方法を提供する。また、本発明は、上記本発明の方法において、前記連結体の塩基配列を決定する工程をさらに含む、複数のmRNAの5'末端領域の塩基配列の解析方法を提供する。さらに、本発明は、請求項1ないし10のいずれか1項に記載の方法において、前記第1工程に代えて、予め得られている完全長cDNAを出発材料として用いる方法を提供する。さらに、本発明は、上記本発明の方法により作製された連結体を提供する。さらに、本発明は、本発明の連結体を含むベクターを提供する。さらに本発明は、本発明の方法により作製された前記連結体から誘導された配列タグを提供する。本発明はさらに、前記連結体から誘導された配列を、複数のRNA試料の内容物を分析するために用いる手段を提供する。さらに本発明は、前記連結体から誘導された配列を、遺伝子の調節及び発現に必要な、ゲノム中の領域の同定に用いる手段を提供する。

【0011】

本発明は、連結体を5'末端のシーケンシングに用いることに限定されるものではなく、5'末端のふとう付加の特定の工程の修飾及び個々に開示するこれらのクローニングは、特定の5'末端の個々のシーケンシングを可能にする。本発明のこのような具体例は、固体マトリックスに特異的に結合されるリンカーを用いる、第1及び第2工程の修飾を包含する。支持体に結合されたcDNAは、次

に、シーケンシング反応を作るために用いられる。

【0012】

このように、本発明は、より一般的には、転写された遺伝子の5'末端に対応する核酸の部分を単離し、それらを、シーケンシングのようなさらに効率的な分析に用いるという思想に関する。

【0013】

【発明の実施の形態】

上記の通り、本発明の方法は、限定されるものではないが、大きく分けて3つの工程を含み、各工程はさらに複数の工程を含む。以下、各工程について説明する。なお、各工程の具体的な実施方法は、下記実施例に詳細に記載されている。

【0014】

第1工程

第1工程は、例えば、試料中のmRNAを鋳型として形成される複数のcDNAであって、該mRNAの5'末端領域に相補的な領域を含むcDNAを選択的に回収する工程である。

【0015】

出発材料としては、所望の細胞や組織から採取した全RNA又はmRNAを用いることができる。全RNAやmRNAの調製方法は周知であり、下記実施例にも詳細に記載されている。他の具体例では、遺伝子の転写部分の5'末端に対応する5'末端核酸を単離するために全長cDNAライブラリーを用いることができる。あるいは、5'末端近傍にクラスIIS酵素部位を有するのであれば、cDNAライブラリー自体を開裂することもできる。

【0016】

第1工程自体は公知の方法により行うことができる。すなわち、完全長cDNAを構築する方法や、少なくともmRNAの5'末端領域に対応する領域を含むcDNA断片の合成方法が知られており、これらの公知の方法のいずれをも採用することができる。好ましい方法として、キャップトラッパー(cap trapper)法(例えばPiero CARNINCI et al., METH

ODS IN ENZYMOLOGY, VOL. 303, pp. 19-44, 1999等) を挙げることができる。以下、このキャップトラッパー法について説明する。もっとも、本発明は、キャップトラッパー法を用いる方法に限定されるも尾ではなく、全長cDNAを豊富化又は選択する他の方法も同様に適用することができる。1つの代替法 (Pelletierら、1995に記載されるように) では、ハイブリッドのRNase処理後に、固定化キャップ結合性タンパク質を用いて全長cDNAを単離する。

【0017】

キャップ選択に代えて、mRNAの5'末端を、BAP (細菌アルカリフェオスマターゼ) のようなフェオスマターゼで脱リン酸化し、次いで脱キャップ化酵素 (タバコアシッドピロフェオスマターゼ) で処理することができる。次いで、元のキャップ構造に代えて、リボヌクレオチド又はデオキシリボヌクレオチドをmRNAの5'末端にRNAリガーゼにより結合することができる (Maruyama及びSugano)。このようにして、例えば、ライゲーション工程の間に、オリゴヌクレオチド/リボヌクレオチド配列上に、cDNA又はRNAの5'末端に、クラスIIs認識部位を置くことができる。このクラスIIs制限酵素は、cDNAを開裂して5'末端タグを生成することができる。

【0018】

ビオチンに代えてキャップ結合性タンパク質 (Pelletier et al., Mol. Cell. Biol. 1995 15:3363-71) や、キャップ構造に特異的に結合する抗体を、前記した選択結合性物質として用いることもできる。

【0019】

あるいは、Gensetにより記載されたように、キャップ構造にオリゴヌクレオチドを科学的に結合することもできる (米国特許6,022,715)。これは、(1) クラスIIs酵素部位を含むオリゴヌクレオチドをキャップに結合すること、(2) 第一鎖合成、次いで、これを第二鎖合成に切り換えること; 第二鎖合成後、cDNAをクラスIIs酵素で開裂して5'タグを作製し、次いで連結体を形成することを可能にする。

【0020】

あるいは、クローンテックにより記載されたキャップスイッチ法（米国特許5, 962, 272）を用いることもできる。核酸を認識することができる物質のための認識部位を有する第一鎖cDNAを、キャップスイッチオリゴヌクレオチドの存在下で作製し、クラスIIs制限酵素部位のような前記認識配列から遠く離れた部位でそれらを開裂する。該キャップスイッチメカニズムは、第一鎖配列を、キャップスイッチオリゴヌクレオチド上に継続させることができる。次いで、第二鎖cDNAを合成し、次いでさらにまたPCR（例えば、SMART（商標）クローンテッククローニングシステムに記載されたような）を行い、最後にクラスIIsで開裂して5'末端タグを生成することができる。

【0021】

他の具体例では、RNAの品質が許す場合には、プライミングによりcDNAを合成し、RNAをキャップ構造まで伸長させることができる。特定の酵素及び反応条件により、ときどき極めて効率的にキャップ部位まで到達することができる（Carninci et al, Biotechniques, 2002）。キャップ選択なしでさえ、キャップ構造に代えて、後で連結体を精製するのに用いられるクラスIIs制限酵素部位を有するオリゴヌクレオチドを結合することができる。

【0022】

キャップトラッパー法では、先ず、RNAを鑄型として、逆転写酵素を用いて第一鎖cDNAを合成する。これは周知の方法により行うことができる。用いるプライマーは、オリゴdTプライマーでもよいし、鑄型RNAがmRNAの場合にはランダムプライマーを用いてもよい。反応溶液中にトレハロースを加えると逆転写酵素が活性化されて逆転写の効率が向上するので好ましい。また、dCTPに代えて5-メチル-dCTPを用いると、いくつかの制限酵素によって切断されなくなるので、工程中、予想外の制限酵素による切断をかなり防止でき、好ましい。また、第一鎖cDNA合成後、CTAB（セチルトリメチルアンモニウムプロミド）等で処理することにより、又はcDNAを精製するより一般的な方法により、タンパク質やペプチドを除去しておくことが好ましい。

【0023】

次に、mRNAのキャップ構造に選択結合性物質を結合する。ここで、「選択結合性物質」とは、ある物質と選択的に結合する物質を意味し、好ましい例としてビオチンを挙げることができるが、これに限定されるものではない。キャップ構造は、mRNAの5'末端に存在する構造であり、転移RNA(tRNA)やリボソームRNA(rRNA)には存在しない。従って、出発材料として全RNAを用いた場合でも、選択結合性物質が結合されるのはmRNAのみである。また、mRNAであっても、5'末端のキャップ構造が切断されてしまっているものには選択結合性物質は結合されない。mRNAのキャップ構造にビオチンを結合させることは、公知の方法により行うことができる。例えば、mRNAをNa₁₀₄のような酸化剤で処理することによりキャップ構造上のジオール基を酸化し、これにビオチンヒドラジド等を反応させることによりキャップ構造にビオチンを結合することができる。あるいは、全長cDNAの作製分野の当業者に知られている他のいずれの方法も、本発明の5'末端を選択的に豊富化するために用いることができる。

【0024】

次に、RNase I処理等により、一本鎖RNAを切断する。あるいは、一本鎖RNAを切断できるがcDNA/RNAハイブリッドを切断できない他のRNase、又は種々の一本鎖RNAを種々の特異性で切断することができるRNaseの混合物も用いることができる。第一鎖cDNAがRNAの5'末端に対応する領域まで伸長されなかったRNA/cDNAハイブリッドは、RNAの5'末端近傍がcDNAとはハイブリダイズできずに一本鎖となっているから、この一本鎖の部分で切断されてしまい、キャップ構造が失われる。従って、この処理により、キャップ構造を維持しているのは、mRNAの5'末端まで完全に伸長したmRNA/cDNAハイブリッドのみとなる。

【0025】

一方、上記選択結合性物質と選択的に結合する、支持体上に固定化された対応選択結合性物質を準備する。なお、本明細書において、「対応選択結合性物質」とは、上記選択結合性物質と選択的に結合する物質を意味し、選択結合性物質が

ビオチンの場合には、アビジン若しくはストレプトアビジン又はビオチン若しくはその誘導体と特異的に結合するこれらの誘導体である。また、支持体としては、特に限定されないが、磁性ビーズを好ましく用いることができ、とりわけ、磁性多孔性ガラスビーズを好ましく用いることができる。ストレプトアビジンを固定化した磁性多孔性ガラスビーズが市販されているので、このような市販のストレプトアビジン固定化磁性多孔性ガラスビーズを好ましく用いることができる。ラテックスビーズ、ラテックス磁性ビーズ、アガロースビーズ、ポリスチレンビーズ、セファロースビーズ又はアルカリのような他の材料も、多孔性ガラスビーズに代えて用いることができる。さらに、本発明は、ビオチニアビジン系を用いるものに限定されるものではなく、ジゴキシゲニンタグをキャップ構造に結合し、ジゴキシゲニンを認識する抗体を固相マトリックスに結合するような、他の結合性物質も用いることができる。

【0026】

次いで、上記したキャップ構造を有するmRNA/cDNAハイブリッドと、支持体に固定化された上記対応選択結合性物質とを反応させ、キャップ構造上の選択結合性物質と、支持体上の対応選択結合性物質とを結合させ、それによってキャップ構造を有するmRNA/cDNAハイブリッドを支持体上に不動化する。支持体が磁性ビーズの場合には、磁力をかけて磁性ビーズを迅速に回収することができる。上記の通り、この段階でキャップ構造を有しているのは、mRNAの5'末端まで完全に伸長したmRNA/cDNAハイブリッドのみであるから、これによりmRNAの5'末端領域に相補的な領域を含むcDNAが選択的に回収され、第1工程が完結する。なお、支持体への非特異吸着を防止するため、この反応を行う前に予め支持体を過剰量のDNAフリーtRNAで処理してプロッキングしておくことが好ましい。表面のプロッキングに公的な他の物質は、全RNA又はオリゴヌクレオチドのような、核酸又は誘導体；例えばウシ血清アルブミンのようなタンパク質；並びに例えばグリコーゲン、デキストラン硫酸、ヘパリン又は他の多糖類のような多糖類である。あるいは、上記した全ての物質の部分を含むハイブリッド分子を用いて非特異的結合部位をマスクすることができる。

【0027】

なお、上記説明では、第1工程をキャップトラップ法で行う場合について説明したが、mRNAの5'末端領域に相補的な領域を含むcDNAを選択的に回収することができる方法であれば、他の種々の方法を用いることもできる。

【0028】

続く第2工程では、mRNAの5'末端領域に相補的な領域と相補的な領域を少なくとも含むcDNA領域を含む断片を選択的に回収する。

【0029】

先ず、支持体に不動化されている第一鎖cDNAを支持体から遊離させる。これは、支持体を例えばNaOHのようなアルカリで処理することにより行うことができる。あるいは、アルカリに代えて、RNaseH（DNAにハイブリダイズしたRNAのみを切断する）による酵素反応を用いることもできる。アルカリ処理により、mRNA/cDNAハイブリッドが支持体から遊離し、さらに、mRNAは分解されて第一鎖cDNAだけが残る。

【0030】

次いで、認識されたDNAを、認識配列の外側で切断する酵素活性を有する物質によって配列特異的に認識されることができる配列を有するリンカー。このような物質の1例が、クラスIIs制限酵素である。

【0031】

この具体例では、クラスIIs制限酵素部位と、3'末端領域にランダムオリゴマー部分とを少なくとも有するリンカーを、上記mRNAの5'末端（すなわちcDNAの3'末端）に対応する、この第一鎖cDNAの末端に連結する。後で5'末端配列タグを連結体にクローニングするために、例えばクラスIIs制限酵素のためい用いられる上記認識部位とは異なる、第2の認識部位をリンカーに導入することが好ましいが必須的ではない。

【0032】

これは、好ましくは、クラスIIs制限酵素部位とランダムオリゴマー部分を有するリンカーを用いた次のような方法（SSL LM (single strand linker ligation method, Y. Shibata

et al., Biotechniques, Vol. 30, No. 6, pp. 1250-1254, (2001)により行うことができる。クラスIIs制限酵素とは、制限酵素認識部位以外の部分で切断が起きる制限酵素群である。1例としてGsuIを挙げられるがこれに限定されるものではない。GsuIで処理すると、一本の鎖が認識部位の16bp下流、他方の鎖が認識部位の14bp下流で切断される。他の適当な例として、認識配列からそれぞれ20塩基及び18塩基離れた部位を切断するMmeIを挙げることができる。ランダムオリゴマー部分は、リンカーの3'末端領域に位置し、塩基数は特に限定されないが5~9、さらには5又は6が適当である。また、クラスIIs制限酵素部位は、切断点がcDNAの中に来るよう、とりわけ、cDNAの比較的5'側（鋳型となったmRNAの3'側）の内部にまで来るよう、上記ランダムオリゴマー部分に近い位置に存在することが好ましい。また、リンカーは、上記ランダムオリゴマー部分が3'側に突出して接着末端を与える二本鎖DNAリンカーであることが好ましい。また、後で容易に回収できるよう、リンカーには、ビオチン等の選択結合性物質を結合しておくことが好ましい。

【0033】

上記第一鎖cDNAを、このようなリンカーと反応させると、リンカーのランダムオリゴマー部分が第一鎖cDNAの3'末端（鋳型となったmRNAの5'末端）領域とハイブリダイズする。次いで、このリンカーをプライマーとして、第一鎖cDNAを鋳型として第二鎖cDNAを合成する。この工程自体は常法により行うことができる。

【0034】

次に、得られた二本鎖cDNAを上記クラスIIs制限酵素で処理する。そうすると、リンカー由来の部分と、cDNAの5'末端（第二鎖cDNAの5'末端）領域由来の部分とから成る二本鎖DNA断片が得られる。例えば、クラスIIs制限酵素としてGsuIを用い、制限酵素部位を上記ランダムオリゴマー部位の直上流に位置するように設計したリンカーを用いた場合、得られたDNA断片は、第2鎖cDNAの5'末端側領域（すなわちmRNAの5'側末端領域）に由来する領域を16bp（ただし相補鎖は14bp）含む。MmeIを用いた

場合には、第二鎖cDNA断片の長さは、それぞれ20および18bpに増加する。

【0035】

次に、このようなDNA断片を選択的に回収する。これは、リンカーに上記のように選択性結合物質（例えばビオチン）が結合されている場合には、対応選択結合性物質（例えばストレプトアビシン）が固定化された支持体（例えば磁性ビーズ）を用いて、第1工程と同様にして行うことができる。このようにして、第一鎖cDNA中の、前記mRNAの5'末端領域に相補的な領域を少なくとも含むcDNA領域を含む断片を選択的に回収する第2工程が完結する。

【0036】

なお、第2工程をSSL LMにより行う場合について説明したが、第2工程はこれに限定されるものではなく、第一鎖cDNAの3'末端領域（鋳型となったmRNAの5'末端領域）を含む断片を選択的に回収できる方法であれば、他の方法も採用することができる。例えば、5'→3'方向に、制御された速度でヌクレオチドを切斷していくエキソヌクレアーゼを用いることによっても可能である。このようなエキソヌクレアーゼで所定の時間第一鎖cDNAを処理することにより、第一鎖cDNAの3'末端領域（鋳型となったmRNAの5'末端領域）を含む一本鎖断片が残る。二本鎖断片のみを分解するヌクレアーゼで処理することにより、目的とする一本鎖断片のみを得ることができる。これを回収してアダプターを連結し、クローニングすることが可能である。

【0037】

続く第3工程では、回収された断片を互いに連結して連結体を形成する。試料中のmRNAは複数存在し、また、リンカーは上記のようにランダムオリゴマー部分で第一鎖cDNAとハイブリダイズするので、試料中に含まれる複数のmRNA由来の複数のcDNA含有断片が上記方法により得られる。第3工程では、これらの複数の断片を連結し、連結体（コンカテマー）を形成する。cDNA断片の連結は、市販のライゲーションキットを用いて常法により行うことができる。連結は、先の段階で用いられた他の認識部位とは異なる制限酵素の認識部位を与える第2のリンカーを導入し、次いで、これで2つのジータグ断片を連結して

連結体にするという方法により、確実に行っていくことができるがこの方法に限定されるものではない。連結する断片の数は、特に限定されず、2以上であればいくらでもよいが、約30個程度が適当である。得られた連結体は、常法により増幅やクローニングすることが好ましいが必ずしも必要ではない。

【0038】

このようにして得られた連結体は、試料中に含まれる複数のmRNAの5'末端領域の塩基配列と同じ塩基配列（ただし、RNA中のウラシルはDNAではチミンになる）を有する領域をそれぞれ含んでいる。リンカー由来の部分も含まれているが、リンカーは塩基配列がわかっているので、連結体の塩基配列を調べれば、リンカー由来の部分とmRNA由来の部分とは明瞭に区別できる。従って、得られた連結体の塩基配列を決定することにより、試料中に含まれる複数のmRNAの5'末端領域の塩基配列を知ることができる。GsuI又はMmeIを用いる好ましい態様では、各mRNAの5'末端領域の最大16塩基又は20塩基の塩基配列を知ることができる。16塩基又は20塩基あれば、統計学的にmRNAをほぼ確実に特定でき、また、新規mRNAか否かも判定できる。なお、連結体の塩基配列を決定すれば、連結体に含まれる上記断片の数（好ましくは20～30個）のmRNAの5'末端領域の塩基配列を知ることができ、効率的に多数のmRNAの5'末端領域を知ることができる。5'末端由来の配列とリンカー由来の配列とを区別するコンピューターソフトを用いることにより、連結体の分析を自動化することができる。

【0039】

5'末端の塩基配列が新規なmRNAが存在していた場合、この領域をフォワード側のプライマーとし、リバース側はオリゴdTをプライマーとしてRT-PCRを行うことにより、新規mRNA由来のcDNAを得ることができる。あるいは、NASBA法等によりmRNAを増幅することも可能である。従って、本発明の方法は、新規な遺伝子のクローニングに用いることができる。同様に、5'末端特異的情報から誘導されるフォワードプライマーは、既存のcDNAライブラリーからのcDNA断片の一部又は全長を増幅するために用いることができる。

【0040】

上記方法では、試料中のmRNA又は全RNAを出発材料として用いたが、予め得られている完全長cDNAライブラリーを用いて第1工程を省略することも可能である。この方法によれば、完全長cDNAライブラリーに含まれる複数のcDNAの5'末端領域（すなわち、該cDNAの鑄型になったmRNAの5'末端領域）の塩基配列情報を上記と同様に効率的に得ることができる。

【0041】

いくつかの具体例においては、転写領域の5'末端から伸長された配列情報を得ることが望まれる。このような伸長された配列は、特定の場合には、タンパク合成の開始部位の同定又はゲノム配列のより良いマッピングを可能にする。上記した通り、本発明は、第二工程において、リンカーをcDNAの5'末端に結合することを含む。このようなリンカーは、特定の核酸断片との結合又は連結のために、連結体から得られる配列を包含する一本鎖オーバーハングを導入することにより修飾することができる。豊富化後に遊離することが可能な支持体にリンカーを結合することにより、DNA断片を豊富化するために、連結後にリンカーを用いることができる。リンカーはさらに、5'末端の伸長された配列情報を得るためのプライマーとして用いることもできる。

【0042】

本発明の方法により得られる連結体の塩基配列を調べることにより、上記したように新規遺伝子のクローニングが可能になるのみならず、試料中の遺伝子の発現プロファイルも調べることができる。さらに、転写開始部位のマッピング、プロモーター利用（ユーセージ）パターンのマッピング、プロモーターのSNPsの解析、発現解析、オルタナティブプロモーターユーセージ及び他のデータとの組合せによる遺伝子ネットワークの創製、断片化ゲノミックDNA中のプロモーター領域の選択的回収（mRNAの5'末端領域と同じ塩基配列を有する断片（ビオチン等を結合）をタグとして用いて断片化ゲノミックDNAとハイブリダイズさせ、ストレプトアビシン固定化磁性ビーズ等を用いてハイブリダイズしたゲノミックDNA断片を回収し、所望によりライブラリー化する）等の種々の用途に用いることができる。

【0043】

あるいは、全長cDNAの混合物を、オリゴヌクレオチドの均一な配列を担持する磁性ビーズに結合し、次いで、SSLMと同様に連結し、第二鎖cDNAを作成し、クラスIIs制限酵素で開裂することにより、連結体の作成を避けることができる。5'末端特異的タグは、ビーズに特異的に固定し、Lyinx治療（米国特許6,352,828; 6,306,597; 6,280,935; 6,265,163; 5,695,934）によってなされるのと同様に特異的シーケンシングのために用いることができる。

【0044】

例えば、オリゴヌクレオチドは、cDNAの5'末端に結合する「ランダム部分I」と、連結産物をに「タグ」を付けることを可能にする、オリゴヌクレオチドのコード部とを含む。cDNAとハイブリダイズしないオリゴヌクレオチドは、エキソヌクレアーゼVIIにより破壊することができる。「解読」オリゴヌクレオチドを用いて配列を選択することができる。ビーズ上のcDNAの特異的なアレーは、次いで、固相表面の1つの位置にそれぞれ整列され、次に並行的にシーケンスを行うことができる。もしあなたが、1個のビーズあたり1個の穴を見るのであれば、あなたは特定のオリゴヌクレオチドを有するビーズのアレイを作ることができる。

【0045】

5'末端の直接シーケンシングのための上記したアプローチを修飾することにより、本発明は、連結体の形態にある5'末端の一般的な解析のため、または5'末端特異的センタKにより豊富化された、個々の5'末端の解析のための異なる手段を提供する。

【0046】

【実施例】

以下、本発明を実施例に基づきより具体的に説明する。もっとも、本発明は下記実施例に限定されるものではない。実施例に岸下実験は、分子生物学分野の当業者により実施することができる。特に断りがない限り、実施例で用いる技術用語、略語及び溶液は、本発明の分野における当業者によって一般的に理解されて

いる意味を有する。このような用語、略語及び溶液の一般的な記述は、Molecular Cloning (Sambrook and Russell, 2001) の一般的試薬のセクションに記載されている。ここに記載される全ての刊行物は、この明細書に組み入れられたものとし、その中の方法及び／又は材料が記載されているものとする。

【0047】

実施例 1

組織からの全RNAの調製

文献では、RNAの調製方法として種々の異なるアプローチが記載されており、それらは当業者に知られている。これらの全てのアプローチは、組織及び細胞を包含する生物学的材料から複数のRNA試料を調製するために用いることができ、それらは本発明に適している。以下に、このような操作の2つを記載する。

緩衝液及び溶液

a) 溶液D： 4M グアニジニウムチオシアネート、25mMクエン酸ナトリウム (pH 7.0), 100mM 2-メルカプトエタノール及び0.5% n-ラウリルーザルコシン

b) RNaseフリーCTAB／尿素溶液：1% CTAB (Sigma)、4M尿素、50mM Tris-HCl (pH 7.0), 1mM EDTA (pH 8.0)

c) Molecular Cloning (Sambrook and Russell, 2001) に記載されたような水平衡化フェノール。

Molecular Cloning (Sambrook and Russell, 2001) に記載されたようなリン酸緩衝液 (PBS)。

5M 塩化ナトリウム

7M 塩化グアニジウム

RNaseフリー蒸留脱イオン水

【0048】

全RNAの調製のためのプロトコール

冷却した皿中で、可能な限り迅速に組織を切り取る。

50ml ファルコン (f a l c o n) チューブ中で組織の体積を測定する。組織の量は、好ましくは、20ml の溶液D当たり0.5~1gである。

2ml の2M酢酸ナトリウム (pH 4.0) を加え、16ml の水平衡化フェノールを加える。ボルテックスミキサーで攪拌する。4ml のクロロホルムを加え、手及びボルテックスミキサーで激しく攪拌する。氷上に15分間静置する。

4°Cで、6000 rpmで30分間遠心する。

水相をピペット (25ml) で新たなチューブに移す。約20ml の水相が得られる。

等体積 (この場合約20ml) のイソプロパノールを加えてRNAを水相から沈殿させ、氷上で1時間静置する。

4°Cで7500 rpmで15分間遠心する。RNAは、遠心により沈殿する。

SCN塩が除去するために、沈殿 (ペレット) を70%エタノールで2回洗浄する。各洗浄の後には7500 rpmで2分間遠心する。

CTABで多糖類を除去する。すなわち、RNAを4ml の水に完全に再懸濁した後、1.3ml の5M NaClを加え、16ml のCTAB／尿素溶液を加えることによりRNAを選択的に沈殿させる。

7500 rpm (9500 x g) で15分間遠心し、水相を捨てる。

RNAペレットを、4ml の7M GuC1中に再懸濁する。

再懸濁したRNAを、8ml のエタノールを加えることにより沈殿させる。-20°Cで1~2時間インキュベートし、4°Cで7500 rpmで15分間遠心する。最後に、ペレットを5ml の70%エタノールで沈殿させる。

7500 rpmで5分間、再度遠心する。

上清を捨てる。

RNAを500~1000μl のRNaseフリーの脱イオン蒸留水 (ddH₂O) に再懸濁する。

【0049】

mRNAの調製

市販のキット、例えばMACS mRNA単離キット (Millipore社製) やpolyA-quick (Stratagene社製) 等を用いることにより

容易に行うことができる。オリゴdTのみをプライマーとして用いる場合には、全RNAを用いることが便利である。リボソームRNA（キャップ構造を有しない）由来のcDNAは、完全長cDNAの選択工程で除外される。pol y-A+RNAは、1~2μg/μlの高濃度に再溶解することが好ましい。

【0050】

cDNAライブラリーからの複数のRNA試料の調製

あるいは、遺伝子の5'末端に対応する複数の核酸を、発現ベクターにクローニングされた既存のcDNAライブラリーから得ることもできる。分子生物学分野の当業者に知られた標準的な方法により、このようなライブラリーから、例えば、T3, T7又はSP6 RNAポリメラーゼを用いたインビトロの転写反応により、RNA転写物を得ることができる。このようなアプローチは、先ずプラスミドDNAを適当な制限エンドヌクレアーゼによって線形化することにより行われる。制限酵素は、センスRNAの転写を可能にするように選択される。ベクター-pFLC III (Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watahiki A, Shibata K, Konno H, Muramatsu M, Hayashizaki Y.) Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis, *Genomics*, 2001 Sep; 77 (1-2): 79-90) 中に得られるライブラリーの場合、ベクターはホーミング(homing)エンドヌクレアーゼI-Ceu I又はPI-Sce Iで切断するうことにより線形化することができ、それによって挿入物の途中での切断を避ける。消化のために、チューブ中で、

プラスミドDNA 100マイクロG

10x緩衝液 40マイクロL

制限酵素 100u

d d H₂O a d 400マイクロL

適当な温度で少なくとも2時間インキュベートし、反応混合物の1μLをアガロースゲル電気泳動によって分析する。消化が完了したら、次のものを添加する。

0. 5M EDTA 8マイクロL

10% SDS 8マイクロL

プロテイナーゼK (10mg/ml) 5マイクロL

試料を500マイクロLのフェノール／クロロホルムで抽出する前に、45℃で15分間インキュベートする。水相を500マイクロLのクロロホルムで2回再抽出する。最後に、線形化DNAを標準的な条件下でイソプロパノール又はエタノールで沈殿させ、50マイクロLのTEに溶解する。

【0051】

インピトロRNA合成

RNAフリー条件下で、チューブ中で以下のものを混合する。

線形化プラスミドDNA	20マイクロG
5x T7又はT3緩衝液	200マイクロL
0. 1M DTT	100マイクロL
2mg/ml BSA	40マイクロL
10mM rNTPs	50マイクロL
T7又はT3 RNAポリメラーゼ	10マイクロL
d d H ₂ O	a d 1000マイクロL

37℃で3～4時間、次のものを添加する前にインキュベートする。

10mM塩化カルシウム	10マイクロL
1U/マイクロL DNase RQ1	5マイクロL

37℃で20分間、次のものを添加する前にインキュベートする。

0. 5M EDTA	10マイクロL
10mg/mlプロテアーゼK	5マイクロL

塩化ナトリウムを終濃度1Mに加える前に、45℃で30分間インキュベートする。標準条件下で、フェノール／クロロホルム抽出、次いで、クロロホルムによる再抽出を行うべきである。RNA転写物は、最終的には、イソプロパノール又

はエタノール沈殿により回収することができる。ペレットを200マイクロLの水又はTEに再懸濁する。RNA転写物の品質は、アガロースゲル電気泳動及び定量により確認すべきである。

【0052】

2. 第1鎖cDNA合成

緩衝液及び溶液

- ・飽和、約80%、トレハロース水溶液、低金属含量
- ・4. 9M高純度ソルビトール（例えばFluka社製cat #85529）
- ・任意的：宝酒造社（Takara）製GC-Taq緩衝液

【0053】

酵素及び緩衝液

- ・RNase H-逆転写酵素Superscript III (Invitrogen) 及び緩衝液並びに他の逆転写酵素

【0054】

核酸及びオリゴヌクレオチド

- ・精製第一鎖プライマー (oligo-dT)。

(用いたプライマーの配列：

5' - GAGAGAGAGAGGATCCTTCTGGAGAGTTTTTT
TTTTTTTTTVN-3')

これに代えて、又はこれに追加してランダムプライマー (dN₆-dN₉) (ただし、Nは任意のヌクレオチド)。

- ・mRNA。2. 5~25μg推奨。

- ・あるいは、全RNA、5~50μg

【0055】

放射性化合物

- ・[α-32P] dGTP

【0056】

プロトコールA：トレハロース/ソルビトールによる促進

第一鎖cDNAを調製するために、三本の異なる0.5mlPCRチューブ（A, B, C）に入っている下記試薬を混合する。

【0057】

チューブA：以下のものを添加。

mRNA 2~2.5μg

又は

全RNA 5~50μg

RNAはエタノールで沈殿させ、この段階で直接再懸濁して使用量を最少化してもよい。

第一鎖プライマー (2μg/μl) 14μg (7μl)

総体積： 22μl

該混合物 (mRNA、プライマー) を65℃で10分間加熱し、mRNAの二次構造を溶解する。

チューブB：以下のものを添加。

5X第一鎖緩衝液 28.6μl

0.1M DTT 11μl

dATP, dTTP, dGTP及び5-メチル-dCTP各10mM

9.3μl

4.9Mソルビトール 55.4μl

飽和トレハロース 23.2μl

RNase H-Superscript II逆転写酵素 (200U/μl)
15.0μl

最終体積：142.5μl

【0058】

温度サイクル：40℃4分間、50℃2分間、56℃60分間

出発物質として全RNAを用いる場合には、温度サイクルは、40℃2分間、-0.1℃/分の冷却速度で35℃へ、50℃2分間、56℃60分間

あるいは、ランダムプライマー (dN₉、Nは任意のヌクレオチド) でcDNAをプライムする場合には、25℃。

【0059】

チューブC：

・ 1~1. 5 μ l の [α -32P] dGTP

【0060】

コールドスタートの場合には次のように操作する。

チューブAとBを氷上で迅速に混合する。

40 μ l のA+B混合物をチューブCに移す。

チューブA+B及びCを上記温度サイクルの第1ステップに40℃で迅速に移し、40℃4分間のアニーリングを行う。

反応を、サーマルサイクラーのセッティングに従って行わせる。

【0061】

ホットスタートの場合には次のように操作する。

チューブA、B及びCをサーマルサイクラーに移す。

サイクリングを開始する。

温度が42℃に達した時、迅速にチューブAとBを混合する。

40 μ l のA+B混合物をチューブCに移す。

反応を、サーマルサイクラーのセッティングに従って行わせる。

【0062】

プロトコールB：GCIートレハロースーソルビトールによる促進

チューブA：最終体積：22 μ l に以下のものを添加。

mRNA 5~25 μ g

(エタノールで沈殿させ、プライマーと共に直接再懸濁)

又は

全RNA 50 μ g 以下 (少量プロトコールの場合)

精製第一鎖cDNAプライマー (2 μ g/ μ l) 14 μ g (7 μ l)

最終体積：22 μ l

チューブB：以下のものを添加。

2X GCI (LA Taq) buffer (Takara) 75 μ l

dATP, dTTP, dGTP, 及び5-メチル-dCTP, 各10 mM

	4 μ l
4. 9 Mソルビトール	20 μ l
飽和トレハロース（約80%）	10 μ l
Superscript II 逆転写酵素 (200U/ μ l)	15 μ l
d d H ₂ O	4 μ l
最終体積：	128 μ l
チューブC：	
· α -32P-dGTP	1. 5 μ l

残りの操作は通常の反応条件に従う。通常、ホットスタートが好ましいプロトコールである。予めサーマルサイクラーに次のサイクルを設定しておく。

42°C 30分間、50°C 10分間、4°C 不定時間

【0063】

次のように操作する。

- 1) チューブA、B及びCをサーマルサイクラーに移す。
- 2) サイクリングを開始する。
- 3) 温度が42°Cに達した時、迅速にチューブAとBを混合する。
- 4) 40 μ lのA+B混合物をチューブCに移す。
- 5) 反応を、サーマルサイクラーのセッティングに従って行わせる。

最後に、終濃度10mMのEDTAを添加して反応を停止する。

次に [α 32P] dGTPの取り込み量を測定し、cDNAの収量を計算する。 [α 32P] dGTPを測定してcDNAの量を算出することは、以下の操作においても処理が的確に進行しているか否か等の監視に有用である。

【0064】

3. 第一鎖cDNAのCTABによる沈殿

緩衝液及び溶液

CTAB溶液は、RNA調製で用いたものと同じ。

緩衝液及び溶液

CATB溶液は実施例1と同じ。

放射能を測定し、ホットcDNAの一部をアルカリゲル電気泳動のために取つ

た後、「ホット」及び「コールド」の両方の合成された第一鎖cDNA（チューブB及びC）を遠心中部に移し、CTAB沈殿を次のようにして行う。

第一鎖のチューブBとCを混合し、この混合物に以下のものを添加する。

3 μ lの0.5M EDTA（終濃度10mM）

2 μ lの10 μ g/ μ lプロテイナーゼK

45℃又は50℃で少なくとも15分間、通常1時間インキュベートする。

第一鎖cDNA反応溶液128~142 μ lに以下のものを添加する。

32 μ lの5M塩化ナトリウム（RNaseフリー）

320 μ lのCTAB-尿素溶液（RNA抽出に用いたものと同じ）

室温で10分間インキュベートする。

15000rpmで10分間遠心する。

上清を除去する。

100 μ lの7M GuCIで注意深く再懸濁する。

250 μ lのエタノールを添加し、氷上又は-20/-80℃で30~60分間静置する。

15000rpmで10分間遠心する。上清を除去する。

次いで、ペレットを800 μ lの80%エタノールで2回洗浄する。各回、cDNAがペレット化される側と反対側の部分に80%エタノールを加え、15000rpmで3分間遠心する。

cDNAを水に再懸濁する（TE又はTris-含有緩衝液を用いない）。46 μ l

【0065】

4. キャップトラッピング並びにキャップの酸化及びビオチン化

緩衝液及び溶液

- 1M酢酸ナトリウム緩衝液、pH4.5

- 1Mクエン酸緩衝液、pH6.0

- Na₂WO₄溶液 > 100mM。新たに調製したものののみを使用。

- SDS 10%

- ビオチン化緩衝液：33mMクエン酸ナトリウム、pH6.0、及び0.3

3% SDS

・クエン酸／SDS緩衝液中 10 mMビオチンヒドラジン ロングアーム (10 ng arm)

(MW=371.51; 3.71 mg/ml = 10 mM)

キャップのビオチン化： (A) mRNAのジオール基の酸化

最終体積 50～55 μl になるように、以下のものを添加する。

再懸濁した cDNA 試料

3. 3 μl の 1M クエン酸ナトリウム緩衝液、pH 4.5

最終濃度 10 mM の、新たに調製された Na1O4 溶液

氷上で 45 分間インキュベートする。

最後に次のようにして cDNA を沈殿させる。

以下の工程を単純化するために、1 μl の 80% グリセロールを添加する。

ボルテックスミキサーで攪拌する。

0.5 μl の 10% SDS, 11 μl の 5M NaCl 及び 61 μl の イソプロパノールを添加する。

-20/-80°C で暗黒化で 30 分間静置する。

15000 rpm で 15 分間遠心する。

上清を除去する。

80% エタノール 500 μl を添加する。

15000 rpm で 2～3 分間遠心する。

上清を捨てる。

工程 12) - 13) を繰り返す。

cDNA を 50 μl の 水中に再懸濁する。

ビオチン化： (B) 酸化ジオール基の誘導

cDNA (50 μl) に、反応緩衝液中の 160 μl の ビオチンヒドラジドロングアーム溶液を加える。反応を 210 μl (最終体積) で行う。

室温 (22～26°C) で一夜、又は 37°C で 3～4 時間インキュベートする。

次に、ビオチン化 cDNA を沈殿させるために以下のものを添加する。

75 μl の 1M クエン酸ナトリウム、pH 6.1

5 μ l の 5M NaCl

750 μ l の 絶対エタノール

氷上で 1 時間又は -80/-20°C で 30 分以上インキュベートする。

サンプルを 15000 rpm で 10 分間遠心する。

沈殿を 70% 又は 80% エタノールで 2 回洗浄し、遠心する。

上清を捨て、洗浄を繰り返す。

cDNA を 175 μ l の TE (1 mM Tris, pH 7.5, 0.1 mM EDTA) に溶解する。

キャップトラッピング及び cDNA の 5' 末端領域の遊離

酵素及び緩衝液

RNase ONE (Promega 社製) 及びその反応緩衝液

cDNA サンプルに以下のものを最終体積 200 μ l になるように加える。

20 μ l の RNase I 緩衝液 (Promega)

第一鎖 cDNA 合成に用いた出発 mRNA 又は全 RNA (少量プロトコールの場合) 1 μ g 当たり 1 単位の RNase I (Promega, 5 又は 10 U/ μ l)。

37°C で 30 分間インキュベートする。

反応を停止するため、サンプルを氷上に置き、以下のものを加える。

4 μ l の 10% SDS 及び

3 μ l の 10 μ g/ μ l プロテイナーゼ K。

45°C で 15 分間インキュベートする。

1:1 Tris-平衡化フェノール：クロロホルムで 1 回抽出し、水相を Microcon-100 にロードする。

水で逆抽出を行い、再度 Microcon-Centricon 100 フィルターにロードする。

Microcon 分離を 1 回行う。

8-b) ペレットを 20 μ l の 0.1 x TE で完全に溶解する。

【0066】

磁性ビーズのプロッキング

材料

・ストレプトアビジン被覆M P G (磁性多孔性ガラス粒子) (米国ニュージャージー州のC P G社製)

緩衝液及び溶液

・結合緩衝液：4. 5 M NaCl, 50 mM EDTA, pH 8. 0

特殊装置

1. 5 ml のチューブを保持するための磁性スタンドが必要である。

核酸の非特異吸着をさらに少なくするために、磁性ビーズをDNA-フリー tRNA (10 mg/ml) と予備インキュベートする。

各ライブラリーについて、500 μl の磁性ビーズ (出発mRNA 25 μg当たり) を100 μg のtRNA と予備インキュベートする。

ときどき攪拌しながら氷上で30分間インキュベートする。

磁性スタンド (3分間) でビーズを分離し、上清を捨てる。

500 μl の結合緩衝液で3回洗浄する。

【0067】

5' 末端cDNA捕捉及び遊離

操作

完全長cDNAを捕捉するために、以下のようにしてRNase I一処理cDNAを混合し、ビーズを洗浄する。

- 1) ビーズを500 μl の洗浄／結合緩衝液に再懸濁する。
- 2) ビオチン化第一鎖cDNAを含むチューブに350 μl のビーズを移す。
- 3) ゆっくりと攪拌した後、50°Cで10分間、チューブを回転させる。
- 4) ビオチン化第一鎖cDNA及び350 μl のビーズを含むチューブに、150 μl のビーズを移す。
- 5) ゆっくりと攪拌した後、50°Cで20分間、チューブを回転させる。

磁性スタンドにより、ビーズを上清から分離する。

ビーズの洗浄

下記緩衝液中でビーズをゆっくりと洗浄して非特異吸着されたcDNAを除去する。洗浄体積は0. 5 ml である。

洗浄／結合溶液で2回

0. 3M NaCl／1mM EDTAで1回

0. 4% SDS／0. 5M NaOAc／20mM Tris-HCl pH 8

5／1mM EDTAで2回

0. 5M NaOAc／10mM Tris-HCl pH 8. 5／1mM EDTAで2回

アルカリ遊離（後述）

アルカリによる完全長cDNAのビーズからの遊離

100μlの50mM NaOH, 5mM EDTAを加える。

短く攪拌し、室温で5分間インキュベートし、ときどき攪拌する。

磁性ビーズを分離し、溶離したcDNAを氷上に移す。

100μlの50mM NaOH, 5mM EDTAによる溶離サイクルを2回

以上繰り返し、80～90%のcDNA（ハンドヘルドモニターで測定されるcpmから計算）をビーズから回収する。

溶離分画をプールする。

5'末端プライマーとなる部位のcDNAへの付加

RNase処理

酵素及び緩衝液

RNase ONE（商品名）及びその緩衝液（Promega）

チューブを氷上に置き、中和後のcDNAとRNA（ビーズからのキャリーオーバー）の自発的な再ハイブリダイゼーションを防止する。

チューブに氷上で1M Tris-HCl, pH 7. 0を加え、迅速に混合する。

1μlのRNase I（10U/μl）を加え、迅速に混合する。

37℃で10分間インキュベートする。

RNase Iを除去するために、cDNAをプロテイナーゼKで処理し、逆抽出を包含するフェノール／クロロホルム抽出を行う。

3μgのグリコーゲンを添加する。cDNAを1サイクロのMicro-100（商品名）で処理する。

プライマーとなる部位を添加する前の分画化

材料

・A m e r s h a m - P h a r m a c i a S - 4 0 0 スピニット (又は同様な他のキット)

緩衝液及び溶液

カラム緩衝液：10 mM Tris, pH 8.0, 1 mM EDTA, 0.1% SDS, 及び100 mM NaCl

SDSを含まないカラム緩衝液：10 mM Tris, pH 8.0, 1 mM EDTA and 100 mM NaCl

S - 4 0 0 スピニットカラムクロマトグラフィー

詳細なプロトコールは市販のキットの添付文書に記載されている。以下はS - 4 0 0 スピニットカラムの操作プロトコールである。

カラムを振盪する。

シールを破り、2 mlのチューブに移す。

3000 rpmで1分間遠心する (+4°C)

cDNA (体積 < 20 μl) を添加する。

cDNAの次に80 μlの水を添加する。

3000 rpmで2分間遠心する。

Micron 100 (商品名) で濃縮するか又はイソプロパノールで沈殿させる。回収率は80%を超えるべきである。

【0068】

6. S S L L M

材料

S - 3 0 0 スピニットカラムクロマトグラフィーキット (A m e r s h a m - P h a r m a c i a)。

緩衝液及び溶液

カラム緩衝液：10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% SDS, 100 mM NaCl

酵素及び緩衝液

TaKaRa DNAリガーゼキットII。

核酸及びオリゴヌクレオチド

ここに記載する実施例では、制限酵素Bgl II, Gsu I and Mme Iが導入される。

しかしながら、本発明は、これらの宣言酵素及びその認識部位を用いるものに限定されるものではない。特にBgl II (認識部位：agatct) は、クローニングにとって適當なあらゆるエンドヌクレアーゼに交換することができる。このような酵素の他の例としては、Asc I (認識部位GGCGCGCC) 及びXba I (認識部位：TCTAGA) を挙げることができる。

Gsu I 認識部位を含む、以下のオリゴヌクレオチドを合成する。

オリゴヌクレオチドBgl-Gsu-GN5：

5' -ビオチン-agagagagaactaggcttaataggtagac
tagatctggagggnnnnn-3'

オリゴヌクレオチドBgl-Gsu-GN6：

5' -ビオチン-agagagagaactaggcttaataggtagac
tagatctggagggnnnnn-3'

オリゴヌクレオチドダウ：

5' P-gaattctcaggactttctatagtgtagcc
aagtctctctc-NH₂ 3'

Mme I 認識部位を含む、以下のオリゴヌクレオチドを合成する。

オリゴヌクレオチドBgl-Mme-GN5：

5' -ビオチン-AGAGAGAGAACTAGGCTTAATAGGTGAC
TAGATCTTCCRACGNNNN-3' ;

オリゴヌクレオチドBgl-Mme-N6：

5' -ビオチン-AGAGAGAGAACTAGGCTTAATAGGTGAC
TAGATCTTCCRACNNNNNN-3' ;

オリゴヌクレオチドBgl-Mme-down：

5' P-GTYGGAGATCTAGTCACCTATTAAGCCTAGTT
CTCTCTCT-NH₂ 3' .

ここで、Pはオリゴヌクレオチドの5'末端がリン酸化されていることを示し、NH₂はアミノ基を付加して非特異連結及びヘアピンプライミングを防止していることを意味している。オリゴヌクレオチドは、第一鎖cDNAの場合と同様に、10%ポリアクリルアミドゲル電気泳動で精製する (Sambrook and Russell, 2001)。オリゴヌクレオチドは、第一鎖cDNAの場合と同様、フェノール／クロロホルム及びクロロホルムで抽出し、2倍容のエタノールで沈殿させる。

【0069】

リンカーの調製

ODをチェックした後、B_g-G_su-GN5、B_g-G_su-N6及び「ダウン」オリゴヌクレオチドを4:1:5の割合で混合する。この場合、DNA量が少なくとも2 μ g/ μ lになるようにする。NaClを100mMの最終濃度に添加する。オリゴヌクレオチドは、65℃で5分間、45℃で5分間、37℃で10分間、25℃で10分間アニールする。

【0070】

第一鎖cDNAの連結

リンカーとcDNAを混合する（最終体積；5 μ l）

65℃で5分間加熱して、一本鎖cDNAの二次構造を融解する。

リンカー及びcDNAを氷上に移す。

TAKARA DNAライゲーションキットからの溶液IIを5 μ l添加する。

キットの溶液Iを10 μ l添加する。

10℃で一夜インキュベートする（少なくとも10時間以上）。

ライゲーション反応の最後に、1 μ lの0.5M EDTA、1 μ lの10% SDS、1 μ lの10mg/mlプロテイナーゼK、10 μ lの水を加え、45℃で15分間インキュベートする。

フェノール／クロロホルム及びクロロホルムで処理し、60 μ lのカラム緩衝液で逆抽出する。

ライゲーション後、過剰のリンカーをS-300スピンカラムクロマトグラフィーで除去する。

1) カラムを数回振り（マトリックスが完全に再溶解するまで）、次いでカラムを直立させる。

2) 上部キャップをはずし、次に底部キャップをはずす。

3) カラムの緩衝液を排出する。

2 ml のカラム緩衝液を加え、重力により2回排出する。

カラムを15ml遠心チューブに入れ、室温でスwingアウトローターで400×gで2分間遠心する。

カラムに100μlの緩衝液を加え、次いで400×gで2分間遠心する。溶離した体積をチェックする。もし、加えた量（100μl）と異なる場合には、溶離体積が添加体積と同じになるまでこの工程を繰り返す。

1. 5mlのチューブをキャップ切斷後、15mlの遠心チューブにセットし、カラムにサンプルを加える。400×gで2分間遠心する。

溶離した画分を別のチューブに回収する。カラムに50μlの緩衝液を加え、遠心及び別のチューブへの回収を繰り返す。

工程6を3～5回繰り返す。溶離した画分を分離して維持する。

回収した画分をシンチレーションカウンターで計測する。通常、最初の2～3画分を混合する（cDNAのcpmの80%）。

NaClを終濃度0.2Mに加える。当量のイソプロパノールを加えることによりcDNAを沈殿させる。

沈殿及び80%冷エタノールでの洗浄を2回行った後、水に再懸濁する。

第二鎖cDNA

工程1 65℃5分

工程2 68℃30分

工程3 72℃10分

工程4 +4℃

第二鎖cDNAのための操作

試験管中で次のものを混合する。

cDNA

6マイクロLのLA-Taqポリメラーゼ緩衝液 (Takara)

6マイクロLの2. 5mM(各) dNTP's (Takara)

0. 5マイクロLの [α 32P] dGTP

第2鎖プログラムを開始した後、チューブをサーマルサイクラーに入る。

チューブに31の5U/マイクロLのLAポリメラーゼ又は他の耐熱性ポリメラーゼ混合物を、試料温度が65°Cの際に、第1工程中に入れる。

迅速にしかし完全に混合する。

サーマルサイクラーの最後に、10mM(最終濃度) EDTAを加えることによって反応を停止し、プロテイナーゼK諸利によって反応をきれいにし、フェノールクロロホルム抽出及びエタノール沈殿を行う。Sambrook and Russell, 2001, Molecular Cloning, CSHL press, NY参照。

【0071】

cDNAの切断

次にcDNAをクラスIIs制限酵素で切断する。

緩衝液(10x) (MBI Fermentas社製)	10 μ l
----------------------------	------------

GsuI (1U/ μ l) (5U/ μ g DNAを使用)	Y μ l
---	-----------

ddH ₂ O	X μ l
--------------------	-----------

最終体積	100 μ l
------	-------------

ここで、Y及びXはcDNAの量に依存する。

- 1) 37°Cで1時間インキュベートする。
- 2) 2 μ lの0.5M EDTAを添加する。
- 3) 65°Cで15分間インキュベートして酵素を不活化する。

磁性ビーズの調製

並行して、CPG-MPG(ストレプトアビジン被覆磁性多孔性ガラス粒子)を調製する。キャップトラッパーの場合と同様にして行うことができる。

- 1) 200 μ lのGPGビーズを準備する。
- 2) 5 μ gのtRNA(20mg/ml)を添加する。これにより、cDNAのビーズへの非特異吸着が減少する。
- 3) 室温で10~20分間又は氷上で30~60分間、ときどき振盪しながら

インキュベートする。

- 4) ビーズを磁性スタンドに移して3分間保持し、水相を捨てる。
- 5) 1M NaCl、10mM EDTAで3回洗浄する。少なくともビーズの初期体積以上の洗浄液を用いる。
- 6) ビーズの初期体積と同体積の1M NaCl、10mM EDTAにビーズを再懸濁する。

以上の操作によりビーズは使用可能である。洗浄工程から1時間以内にビーズを使用する。

【0072】

7. 遊離したcDNAタグ

- 1) ビーズとGsuI消化サンプルとを混合する。
- 2) ときどきゆっくり混合しながら室温で15分間インキュベートする。
- 3) 磁性ラックに3分間立てる。
- 4) 上清を回収する。
- 5) 1xBSA（ウシ血清アルブミン）を含む500μlの1xB&Wで4回灌ぐ。
- 6) 200μlの1xリガーゼ緩衝液（NEB）で2回洗浄する。

【0073】

8. 結合されたcDNAへのリンカーの連結：IIリンカーライゲーション

- 1) 20μlのL_oTE（1mM Tris, pH7.5, 及び0.1mM EDTA）にリンカーIIを加える（0.4μg/μl）。
- 2) チューブを65℃で5分間加熱し、室温で15分間静置する。
- 3) TaKaRaライゲーションキットIIの溶液IIを25μl、溶液Iを50μl加える。
- 4) 16℃で一夜インキュベートする。
- 5) ライゲーション後、500μlの1xBSA含有1xB&W緩衝液で4回洗浄する。
- 6) 200μlの1xB&W緩衝液で1回、200μlの1xBSA含有1xBgII緩衝液で2回洗浄する。

【0074】

9. タグ化酵素を用いたcDNAタグの遊離

サンプルに以下のものを添加する。

L o T E X μ l

10 x 緩衝液 10 μ l

B g I I I Y μ l

全量が 100 μ l になるように設定

- 1) ときどきゆっくり混合しながら 37°C で 1 時間インキュベートする。
- 2) 磁石上に置き、上清を回収して新たなチューブに入れる。上清は、遊離された 5' 末端断片を含む。
- 3) L o T E で体積を 200 μ l に増やす。

200 μ l のサンプル（リンカーをタグ付加された 5' 末端断片）に以下のものを加える。

- 1) 133 μ l の 7.5M NH₄OAc
- 2) 3 μ l の 1 μ g / μ l グリコーゲン
- 3) 340 μ l の イソプロパノール

-20°C / -80°C で少なくとも 30 分間インキュベートする。微量遠心機で 4°C で 15000 rpm で 20 分間遠心する。上清を除去する。ペレットを 70% 又は 80% エタノールで 2 回洗浄する。15000 rpm で 3 回遠心し、エタノール洗浄液を除去する。最後に、10 μ l の L o T E 中に再懸濁する。

【0075】

10. タグの連結によるジタグ (d i t a g) の形成

cDNA の 5' 末端断片を連結してジタグを形成する。宝酒造社製のライゲーションキット等を用いてライゲーションを促進することができる。

- 1) T a K a R a ライゲーションキット II の溶液 II を 10 μ l 、溶液 I を 20 μ l 加える。
- 2) 16°C で一夜インキュベートする。
- 3) 10 μ l の ddH₂O、1 μ l の 0.5M EDTA、1 μ l の 10% SDS 及び 1 μ l の 10 μ g / μ l プロテイナーゼ K を添加する。

- 4) 45℃で15分間インキュベートする。
- 5) 1:1 Tris-平衡化フェノール：クロロホルムで1回抽出し、水相を回収する。

フェノールークロロホルム及びクロロホルム処理後、逆抽出する。フェノールークロロホルム抽出は、通常の有機物抽出 (Sambrook and Russell, Molecular Cloning, 2001参照) 又はcDNA処理について上記した通りに行うことができる。

- 6) G-50スピンカラム（サイズ排除）で最小のcDNA断片を除去する。
- 7) 5μgのグリコーゲンをキャリアとして加え、イソプロパノールで沈殿させる。

100μlのサンプル

67μlの7.5M NH₄OAc

5μlのグリコーゲン

180μlのイソプロパノール

- 8) 4℃で20分間遠心する。

- 9) 80%又は70%エタノールで洗浄し、遠心し、エタノールを除去する。

なお、cDNAのクラスIIs制限酵素での切断及びそれ以降の上記工程は、次の方法によっても行うことができる。

cDNAのクラスIIs制限酵素での切断

緩衝液 (x 10) 5μl

GsuI (1U/ml) (5U/μg使用) yμl

dH₂O xμl

最終容量 (5+y+x) = 50μl

- 1) 30℃で1時間インキュベートする。
- 2) 1μlの0.5M EDTAを加える。
- 3) 65℃で1分間インキュベートして酵素を失活させる。

ビーズの準備、キャップトラッパーと同様

- 1) 200μlのビーズに2.5μlのtRNA (20μg/μl) を加える

。

- 2) 氷上で30分以上、時々混ぜながらインキュベートする。
- 3) スタンドに移し、3分間静置後、水相を除く。
- 4) 洗浄液 (1M NaCl / 10mM EDTA) 200 μl で洗浄を3回する。
- 5) 最初に用意したビーズと同じ要領の洗浄液で懸濁する。
- 1) ビーズ懸濁液と、タイプ I I s 制限酵素処理したサンプルを混ぜる。
- 2) 時々かるく混ぜながら15分間、室温でインキュベートする。
- 3) スタンドに立て、2~3分間置き、水相（上清）を取る。
- 4) 1x B&W液で3回洗浄する。
- 5) L o T Eで1回洗浄する。
- 6) L o T E 30 μl に懸濁する。
- 1) 30 μl に懸濁したサンプルに0.4 μg/μl のリンカー I I を5 μl 加える。
- 2) 50℃、2分間インキュベートする。
- 3) 室温で15分間放置する。
- 4) x 10 の緩衝液 4 μl 加える。
- 5) T 4 DNAリガーゼ 1 μl を加える。
- 6) 16℃で2時間、時々混ぜながらインキュベートする。
- 7) 1x B&W液で3回洗浄し、タグ化している酵素の緩衝液で2回洗浄する。
-

x 10 緩衝液 10 μl

L o T E x μl

酵素 y μl

最終容量 (10 + x + y) = 100 μl

- 1) 混ぜながら37℃で1時間インキュベートする。
- 2) スタンドに2~3分間置き、上清を回収する。
- 3) L o T Eで洗浄して容量を200 μl にする。
- 4) 133 μl 7.5M NH₄OAc、3.5 μl 1 μg/ml グリコーゲン、340 μl イソプロパノールを加える。

−20℃/−80℃で30分間以上インキュベートし、15000 r p m、20分間遠心、ペレットは70%又は80%エタノールで洗浄後、4 μ l L o T E に懸濁する。

- 1) x 10 ライゲーション緩衝液 0. 5 μ l
- 2) T 4 DNAリガーゼ 0. 5 μ l
- 3) 16℃で一晩インキュベートする。

ジタグの確認

- 1) 上記のサンプルの一部を使ってジタグの確認を行う。

10 x E x T a q b u f f e r	5 μ l
D M S O	3 μ l
2. 5 Mm d N T P s	1 6 μ l
3 5 0 n g / μ l プライマー	2 μ l
c D N A	x μ l
E x T a q	1 μ l
d H 2 O	y μ l
全容量	(5 + 3 + 1 6 + 2 + 1 + z + y) = 5 0 μ l

サイクル： 95℃、2分間の後、95℃、30秒、55℃、1分間、72℃、1分間のサイクルを28回繰り返し、最後に72℃、5分間

- 2) 反応が終わったら1部をアクリルアミドゲルで確認

ジタグの精製

- 1) 確認できたら残りのサンプルもスケールアップして反応させる。
- 2) 反応後、アクリルアミドゲルで精製を行う。

【0076】

12. アンカー酵素によるc D N Aの切斷

- 1) サンプルを5 μ l のL o T E に再懸濁し、以下のものを以下の順序で加える。

i) L o T E	X μ l
ii) 10 x E c o R I 制限緩衝液	5 μ l
iii) E c o R I	Y μ l

総体積を 50 μ l にする。

- 2) 37 °Cで1時間インキュベートする。
- 3) 1 μ l の 0.5M EDTA, 1 μ l の 10% SDS 及び 1 μ l の 10 μ g/ μ l プロテイナーゼ K を加える。
- 4) 45 °Cで15分間インキュベートする。
- 5) 1:1 Tris-平衡化フェノール：クロロホルムで1回抽出し、水相を回収する。

フェノールークロロホルム及びクロロホルム処理後、逆抽出する（上記参照）

。

- 6) 5 μ g のグリコーゲンをキャリアとして加え、イソプロパノールで沈殿させる。

100 μ l のサンプル

67 μ l の 7.5M NH₄OAc

5 μ l のグリコーゲン

180 μ l のイソプロパノール

- 8) 4 °Cで20分間遠心する。

- 9) 80%又は70%エタノールで洗浄し、遠心し、各回ともエタノールを除去する。

【0077】

1.1. ジタグの連結によるコンカテマーの形成

- 1) 5 μ l の L o T E に再懸濁する。
- 2) TaKaRa ライゲーションキット II の溶液 II を 5 μ l 、溶液 I を 10 μ l 加える。
- 3) 16 °Cで1.5時間インキュベートする。
- 4) 1 μ l の 0.5M EDTA、1 μ l の 10% SDS、1 μ l の 10 μ g/ μ l プロテイナーゼ K を加える。
- 5) 45 °Cで15分間インキュベートする。
- 6) 1:1 Tris-平衡化フェノール：クロロホルムで1回抽出し、水相を回収する。

フェノールークロロホルム及びクロロホルム処理後、逆抽出する（上記参照）。

7) $5 \mu\text{g}$ のグリコーゲンをキャリアとして加え、イソプロパノールで沈殿させる。

100 μl のサンプル

67 μl の7.5M NH₄OAc

5 μl のグリコーゲン

180 μl のイソプロパノール

8) 4°Cで20分間遠心する。

9) 80%又は70%エタノールで洗浄し、遠心し、各回ともエタノールを除去する。

10) 5 μl のddH₂Oに再溶解する。

【0078】

実施例2

上記で得られた連結体は、pBluescript II KS+ (Stratagene) のようなクローニングベクターにさらに連結される。多数のクローニングベクターがこの分野において公知であり、それらは本発明に用いることができる。

標準的連結

3倍過剰の連結体と、EcoRIで線形化した、体積5マイクロLの適当なベクター100ngとを混合する。次に、DNAライゲーションキットVer2 (Takara) の溶液Iの5マイクロLを、挿入物/ベクター混合物と混合する。チューブを16°Cで12~6時間インキュベートする。

【0079】

形質転換：

ライゲーション溶液から塩を除去するために、2マイクロGのグリコーゲン (Roche)、20mMの塩化ナトリウム及び80%エタノールを添加した後、DNAを沈殿させる。DNAペレットを、150マイクロLの80%エタノールで2回洗浄し、次いで、10マイクロLの水に溶解する。脱塩したライゲーション

ン溶液1マイクロLを用い、Cell-Porator等を用い、ElectroMAX™ DH10B™細胞(Invitrogen)を形質転換する(製造者のマニュアルに記載された形質転換方法に従ったBiometrer。形質転換された細菌を選択培地上に置床し、一夜増殖させる。陽性クローンをこれらのプレートから単離し、連結体のさらなる特徴づけを行う。

【0080】

実施例3：連結体のシーケンシング

連結体のシーケンシングは、クローニングのフランкиング領域中のネスティッドプライマーを用い、BigDye Terminator Cycle Sequencing Ready Reaction Kit v2.0 (Applied Biosystems) 及びABI3700 (Applied Biosystems) シーケンサーを用い、製造者の製品説明書に従って行うことができる。連結体は、両端からシーケンスして、全配列をカバーする。

【0081】

実施例4：5'末端配列タグの同定

連結体から得られた配列は、図5に示すようにジタグの構造により特徴付けられる。クローニング工程において用いられた制限酵素の認識部位を有する規定された領域は、特定の配列タグの5'末端のそれぞれに隣接する。したがって、5'末端特異配列タグは、主導の配列解析により、又は適当なコンピュータープログラムを用いる自動的方法により同定することができる。個々の5'末端特異配列タグは、コンピューターファイル又はデータベースシステムに貯えることができる。

【0082】

実施例5：5'末端配列タグの特徴づけ

5'末端特異的配列タグは、Genetics Accelrys Inc. のComputer Group (GCG) package (<http://www.accelrys.com/>) で利用可能な、配列アラインメントのための、例えばNCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) , FASTAのような標準的なソフトウェアソ

リューションによりその同一性を解析することができる。このようなソフトウェアソリューションにより、5'末端特異配列タグを互いに整列させ、ユニーク又は希少なタグを同定することができる。これらはさらに、データベース検索、5'末端配列データベースの構築、5'末端配列データベースを用いた遺伝子の同定、に用いることができる。

5'末端特異タグを用いた、GenBankにおけるBLAST検索の一例を以下に示す。16bpのタグ(5' -ACC TCC CTC CGC GGA G)がヒトTGF-β1(JBC 264 (1989) 402-408)の5'末端から誘導される。

Query= (16 letters) (ACCTCCCTCCCCGGAG)

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)

1,205,903 sequences; 5,297,768,116 total letters

Score E

Sequences producing significant alignments: (bits)

Value

gi|10863872|ref|NM_000660.1| Homo sapiens transforming grow... 3

2 1.1

gi|18590091|ref|XM_085882.1| Homo sapiens similar to transf... 32

1.1

gi|11424057|ref|XM_008912.1| Homo sapiens transforming grow... 32

1.1

gi|7684381|gb|AC011462.4|AC011462 Homo sapiens chromosome 1...

32 1.1

gi|15027087|emb|AL389894.4|LMFLCHR4A Leishmania major Fried...

32 1.1

gi|1943914|gb|U70540.1|LMU70540 Leishmania mexicana amazone...

32 1.1

gi|37097|emb|X05839.1|HSTGFBG1 Human transforming growth fa...

32 1.1

gi|37092|emb|X02812.1|HSTGFB1 Human mRNA for transforming g...

32 1.1

gi|340526|gb|J04431.1|HUMTGFB1PR Homo sapiens transforming ...

32 1.1

gi|18858696|ref|NM_131728.1| Danio rerio forkhead box C1a (... 3

Q 4.2

gi|12004937|gb|AF219949.1|AF219949 Danio rerio forkhead tra...

30 4.2

gi|193604|gb|M13366.1|MUSGPDX Mouse glycerophosphate dehydr...

30 4.2

gi|193601|gb|M25558.1|MUSGPD Mouse glycerol-3-phosphate deh...

30 4.2

gi|63465|emb|V00414.1|GGHI01 Gallus gallus mRNA coding for ...

30 4.2

gi|63444|emb|X13894.1|GGH2AF Chicken histone H2A.F gene

30 4.2

Alignments

>gi|10863872|ref|NM_000660.1| Homo sapiens transforming growth factor, beta 1

(Camurati-Engelmann disease) (TGFB1), mRNA

Length = 2745

Score = 32.2 bits (16), Expect = 1.1

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||:|||||||

Sbjct: 1 acctccctccgcggag 16

>gi|18590091|ref|XM_085882.1| Homo sapiens similar to transforming growth factor, beta 1 (H.

sapiens) (LOC147760), mRNA

Length = 697

Score = 32.2 bits (16), Expect = 1.1

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 7 acctccctccgcggag 22

>gi|11424057|ref|XM_008912.1| Homo sapiens transforming growth factor, beta 1 (TGFB1), mRNA

Length = 2741

Score = 32.2 bits (16), Expect = 1.1

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 1 acctccctccgcggag 16

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)

Posted date: Apr 9, 2002 10:59 AM

Number of letters in database: 1,002,800,820

Number of sequences in database: 1,205,903

Lambda K H

1.37 0.711 1.31

Gapped

Lambda K H

1.37 0.711 1.31

Matrix: blastn matrix:1 -3

Gap Penalties: Existence: 5, Extension: 2

Number of Hits to DB: 6901

Number of Sequences: 1205903

Number of extensions: 6901

Number of successful extensions: 1479

Number of sequences better than 10.0: 16

length of query: 16

length of database: 5,297,768,116

effective HSP length: 15

effective length of query: 1

effective length of database: 5,279,679,571

effective search space: 5279679571

effective search space used: 5279679571

T: 0

A: 30

X1: 6 (11.9 bits)

X2: 15 (29.7 bits)

S1: 12 (24.3 bits)

S2: 15 (30.2 bits)

Top of Form

1: NM_000660. Homo sapiens tran... [gi: [Related Sequences](#), [OMIM](#), [Protein](#), [PubMed](#),
10863872] [my](#), [UniSTS](#), [LinkOut](#)

LOCUS NM_000660 2745 bp mRNA linear PRI 13-F

EB-2002

DEFINITION Homo sapiens transforming growth factor, beta 1 (Camurati-Engelmann

disease) (TGFB1), mRNA.

ACCESSION NM_000660

VERSION NM_000660.1 GI:10863872

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2745)

AUTHORS Derynck, R., Jarrett, J. A., Chen, E. Y., Eaton, D. H., Bell, J. R., Assoian, R. K., Roberts, A. B., Sporn, M. B. and Goeddel, D. V.

TITLE Human transforming growth factor-beta complementary DNA sequence

and expression in normal and transformed cells

JOURNAL Nature 316 (6030), 701-705 (1985)

MEDLINE 85296301

REFERENCE 2 (bases 1 to 2745)

AUTHORS Sporn, M. B., Roberts, A. B., Wakefield, L. M. and Assoian, R. K.

TITLE Transforming growth factor-beta: biological function and chemical structure

JOURNAL Science 233 (4763), 532-534 (1986)

MEDLINE 86261803

PUBMED 3487831

REFERENCE 3 (bases 1 to 2745)

AUTHORS Chang, N. S., Mattison, J., Cao, H., Pratt, N., Zhao, Y. and Lee, C

TITLE Cloning and characterization of a novel transforming growth factor- β -induced TIAF1 protein that inhibits tumor necrosis

factor cytotoxicity

JOURNAL Biochem. Biophys. Res. Commun. 253 (3), 743-749 (1998)

MEDLINE 99119079

PUBMED 9918798

REFERENCE 4 (bases 1 to 2745)

AUTHORS Ghadami, M., Makita, Y., Yoshida, K., Nishimura, G., Fukushima, Y

Wakui, K., Ikegawa, S., Yamada, K., Kondo, S., Niikawa, N. and Tomita, H.

TITLE Genetic mapping of the Camurati-Engelmann disease locus to chromosome 19q13.1-q13.3

JOURNAL Am. J. Hum. Genet. 66 (1), 143-147 (2000)

MEDLINE 20100617

PUBMED 10631145

REFERENCE 5 (bases 1 to 2745)

AUTHORS Vaughn, S. P., Broussard, S., Hall, C. R., Scott, A., Blanton, S. H.

Milunsky, J. M. and Hecht, J. T.

TITLE Confirmation of the mapping of the Camurati-Engelmann locus

to

19q13. 2 and refinement to a 3.2-cM region

JOURNAL Genomics 66 (1), 119-121 (2000)

MEDLINE 20304762

PUBMED 10843814

REFERENCE 6 (bases 1 to 2745)

AUTHORS Lim, J. M., Kim, J. A., Lee, J. H. and Joo, C. K.

TITLE Downregulated expression of integrin alpha6 by transforming growth

factor-beta(1) on lens epithelial cells in vitro

JOURNAL Biochem. Biophys. Res. Commun. 284 (1), 33-41 (2001)

MEDLINE 21268957

PUBMED 11374867

COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final

NCBI review. The reference sequence was derived from X02812

1.

FEATURES Location/Qualifiers

source 1..2745
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19q13. 1"

gene 1..2745
 /gene="TGFB1"
 /note="TGFB; DPD1; CED"
 /db_xref="LocusID:7040"
 /db_xref="MIM:190180"

misc feature 37..113
 /note="pot. hairpin loops-forming region"

variation 72
 /allele="-"
 /allele="C"
 /db_xref="dbSNP:1800999"

variation 79
 /allele="-"
 /allele="C"
 /db_xref="dbSNP:1799753"

CDS 842..2017
 /gene="TGFB1"
 /note="transforming growth factor, beta 1; diaphyse
 al
 dysplasia 1, progressive (Camurati-Engelmann diseas
 e)"
 /codon_start=1
 /db_xref="LocusID:7040"
 /db_xref="MIM:190180"
 /product="transforming growth factor, beta 1
 (Camurati-Engelmann disease)"
 /protein_id="NP_000651.1"
 /db_xref="GI:10863873"
 /translation="MPPSGLRLLPLLLPLLWLLVLTPGPPAAGLSTCKTID

 MELVKRK
 RIEAIRGQIILSKI.RIASPPSQGEVPPGPLPEAVLALYNSTRDRVAGESAEPEPEPEAD
 YYAKEVTRVLMVETHNEIYDKFKQSTHSIYMFFNTSELREAVPEPVLI.SRAELRLR
 LKLKVEQHVELYQKYSNNSWRYLSNRLLAPSDSPEWLSFDVTGVRQWLSRGGEIEGF
 RLSAHCSCDSRDNTLQVDINGFTTGRRGDLATIHMNRPFLLLMATPLERAQHLQSSR
 HRRALDTNYCFSSTEKNCCVRQLYIDFRKDI.GWKWIHEPKGYHANFCLGPCPYIWSLD
 TQYSKVLALYNQHNPGASAAPCCVPQALEPLPIVYYGRKPKEQLSNMIVRSCKCS"

misc_feature 863..910
/note="pot. core sequence of signal peptide (aa -27
2 to
-257)"

variation 870
/allele="C"
/allele="T"
/db_xref="dbSNP:1982073"

variation 915
/allele="C"
/allele="G"
/db_xref="dbSNP:1800471"

misc_feature 938..1600
/note="TGFb_propeptide; Region: TGF-beta propeptide"

misc_feature 953
/note="pot. altern. translation start site"

misc_feature 1035..1043
/note="put. glycosylation site"

misc_feature 1247..1255
/note="put. glycosylation site"

misc_feature 1370..1378
/note="put. glycosylation site"

variation 1632
/allele="C"
/allele="T"
/db_xref="dbSNP:1800472"

mat_peptide 1679..2014
/product="mature TGF-beta (aa 1-112)"

<u>misc_feature</u>	1715..2014
	/note="TGF-beta; Region: Transforming growth factor beta
	like domain"
<u>misc_feature</u>	1721..2014
	/note="TGFB; Region: Transforming growth factor-beta
a	(TGF-beta) family"
<u>misc_feature</u>	2018..2096
	/note="GC-rich region"
<u>promoter</u>	2097..2103
	/note="TATA-box-like region"
<u>misc_feature</u>	2517..2522
	/note="put. polyadenylation signal"
<u>polyA_site</u>	2539
	/note="polyadenylation site"

BASE COUNT 527 a 938 c 801 g 479 t

ORIGIN

1 acccccc ccggaggcaggc cagacagcga gggccccggc cgggggcagg ggggacg
ccc
61 cgtccggggc accccccccg gctctgagcc gcggcgccccgggg ccggcctcgg cccggag
cgg
121 aggaaggagt cgccgaggag cagcctgagg ccccaagatgc tgagacgagc cgccgcc
gcc
181 cccgccactg cggggaggag ggggaggagg agcgggagga gggacgagct ggtcggg
aga
241 agagaaaaaa aactttttagt actttccgt tgccgctggg agccggaggc gcggggaa
cct
301 cttggcgcgca cgctgccccg cgaggaggca ggacttgggg accccagacc gcctccc

ttt

361 gcccggggg acgcttgctc cctccctgcc ccctacacgg cgtccctcag gcccgg

cat

421 tccggaccag ccctcgggag tcgcccaccc ggcctcccgaaagacttt cccaga

cct

481 cgggcgcacc ccctgcacgc cgccttcata cccggctgt ctccctgagcc cccgcgc

atc

541 ctagaccctt tctccctccag gagacggatc tctctccgac ctgccacaga tccccta

ttc

601 aagaccaccc accttctggt accagatcgc gcccacatcg gttttccg tgggata

ctg

661 agacacccccc ggtccaagcc tcccctccac cactgcgc cttctccctga ggagcct

cag

721 cttccctcg aggccctcct acctttgcc gggagaccc cagccctgc agggcgc

ggg

781 cctccccacc acaccagccc tgttcgcgct ctcggcagtgc cggggggcgc ccgcctc

ccc

841 catgcgcgc tccgggctgc ggctgctgcc gctgctgcta ccgctgctgt ggctact

ggt

901 gctgacgcct ggccgcggg ccgcggact atccacatgc aagactatcg acatgga

gct

961 ggtgaagcgg aagcgcatcg aggccatcg cggccagatc ctgtccaagc tgccgc

cgc

1021 cagccccccg agccaggggg aggtgccgc cggcccgctg cccgaggccg tgctcgc

cct

1081 gtacaacagc acccgcgacc gggtggccgg ggagagtgc aaccggagc ccgagcc

tga

1141 ggccgactac tacgccaagg aggtcacccg cgtgctaattt gtggaaaccc acaacga

aat

1201 ctatgacaag ttcaaggaga gtacacacag catatatatg ttcttcaaca catcaga
gct

1261 ccgagaagcg gtacctgaac ccgtgttgct ctccggca gagctgcgtc tgctgag
gag

1321 gctcaagtta aaagtggagc agcacgtgga gctgtaccag aaatacagca acaattc
ctg

1381 gcgataacctc agcaaccggc tgctggcacc cagcgactcg ccagagtgg tatttt
tga

1441 tgtcaccgga gttgtgcggc agtggttgag ccgtggaggg gaaattgagg gctttcg
cct

1501 tagcgccac tgctcctgtg acagcaggaa taacacactg caagtggaca tcaacgg
gtt

1561 cactaccggc cgccgaggtg acctggccac cattcatggc atgaaccggc ctttcct
gct

1621 tctcatggcc accccgtgg agagggccca gcatctgcaa agctccggc accggcg
agc

1681 cctggacacc aactatgtc tcaagtcac ggagaagaac tgctgcgtgc ggcagct
gta

1741 cattgacttc cgcaaggacc tcggctggaa gtggatccac gagccaaagg gctacca
lgc

1801 caactctgc ctcggccct gcccctacat ttggagccctg gacacgcagt acagcaa
ggt

1861 cctggccctg tacaaccagc ataaccggg cgccitggcg ggcgcgtgct gcgtgcc
gca

1921 ggcgtggag ccgtgcccc tcgttacta cgtggccgc aagccaaagg tggagca
gct

1981 gtccaaatg atcgtgcgt cctgcaagtg cagctgaggt cccggccgc cccggcc
cgc

2041 cccggcaggc ccggccccac cccggccgc cccggctgcc ttggccatgg gggctgt

att

2101 taaggacacc gtgccccaaag cccacctggg gccccattaa agatggagag aggactg

cgg

2161 atctctgtgt cattgggcgc ctgcctgggg tctccatccc tgacgttccc ccactcc

cac

2221 tccctctctc tccctctctg ctcctcctg cctgtctgca ctattccctt gcccgcc

atc

2281 aaggcacagg ggaccagtgg ggaacactac tgttagttaga tctatttatt gagcacc

ttg

2341 ggcactgttg aagtgcctta catatatgaa ctcattcgt caccatagca acactct

gag

2401 atggcaggga ctctgataaac acccattttt aagggttgggg aaacaagccc agagagg

tta

2461 agggaggagt tcctgccccat caggaacctg cttagtggg ggatagtgaa gaagaca

ata

2521 aaagatagta gttcaggcca ggcggggtgc tcacgcctgt aatcctagca ctllgg

gag

2581 gcagagatgg gaggatactt gaatccaggc attttagacc agcctggta acatagt

gag

2641 accctatctc tacaaaacac tttttttttt tgtacacctg tggcccagc tactctg

gag

2701 gctaagggtgg gaggatcaact tgatcctggg aggtcaaggc tgcag

//

Bottom of Form

Revised: October 24, 2001.

Query= (16 letters)

Database: GenBank Human EST entries

4,280,058 sequences; 2,114,234,064 total letters

Score E

Sequences producing significant alignments: (bits)

Value

gi|19365764|gb|BM915385.1|BM915385 AGENCOURT_6701642 NIH_MG...

32 0.41

gi|19353768|gb|BM903897.1|BM903897 AGENCOURT_6696012 NIH_MG...

32 0.41

gi|18807810|gb|BM562052.1|BM562052 AGENCOURT_6562015 NIH_MG...

32 0.41

gi|18791603|gb|BM553137.1|BM553137 AGENCOURT_6572574 NIH_MG...

32 0.41gi|16171065|gb|BI908151.1|BI908151 603067456F1 NIH_MGC_118 ... 32 0.41gi|15759271|gb|BI767693.1|BI767693 603060648F1 NIH_MGC_122 ... 32 0.41gi|15343643|gb|BI518851.1|BI518851 603061760F1 NIH_MGC_118 ... 32 0.41gi|14309343|gb|BG899094.1|BG899094 HOA21-1-G9 HOA (Human Os... 332 0.41gi|13662542|gb|BG611171.1|BG611171 602612144F1 NIH_MGC_60 H... 332 0.41gi|12609210|gb|BG115704.1|BG115704 602317174F1 NIH_MGC_88 H... 332 0.41gi|12101282|gb|BF796228.1|BF796228 602258513F1 NIH_MGC_85 H... 332 0.41gi|11152079|gb|BF238160.1|BF238160 601811886F1 NIH_MGC_48 H... 332 0.41gi|11100313|gb|BF206727.1|BF206727 601871105F1 NIH_MGC_19 H... 332 0.41

gi|11100272|gb|BF206686.1|BF206686 601871051F1 NIH_MGC_19 H...

32 0.41

gi|16775383|gb|BM046103.1|BM046103 603625849F1 NIH_MGC_40 H...

30 1.6

gi|19739174|gb|BQ014273.1|BQ014273 UI-H-ED1-axs-h-21-0-UI. s...

28 6.4

gi|19378603|gb|BM928224.1|BM928224 AGENCOURT_6699855 NIH_MG...

28 6.4

gi|19367808|gb|BM917429.1|BM917429 AGENCOURT_6606724 NIH_MG...

28 6.4

gi|19364214|gb|BM913835.1|BM913835 AGENCOURT_6612786 NIH_MG...

28 6.4

gi|19361343|gb|BM910964.1|BM910964 AGENCOURT_6615957 NIH_MG...

28 6.4

gi|18505954|gb|BM456914.1|BM456914 AGENCOURT_6404253 NIH_MG...

28 6.4

gi|18499709|gb|BM450669.1|BM450669 AGENCOURT_6394717 NIH_MG...

28 6.4

gi|16000196|gb|BI859449.1|BI859449 603388188F1 NIH_MGC_87 H... 2

8 6.4

gi|15928460|gb|BI818193.1|BI818193 603032663F1 NIH_MGC_115 ... 2

8 6.4

gi|15431547|gb|BI544235.1|BI544235 603241605F1 NIH_MGC_95 H... 2

8 6.4

gi|15345229|gb|BI520437.1|BI520437 603071622F1 NIH_MGC_119 ... 2

8 6.4

gi|14440373|gb|BI033747.1|BI033747 PM3-NN0223-220201-014-h0... 2

8 6.4

gi|14426676|gb|BI020046.1|BI020046 CM3-MT0291-110101-622-f0... 2

8 6.4

gi | 14081325 | gb | BG770672.1 | BG770672 602734012F1 NIH_MGC_49 H...

28 6.4

gi | 13546630 | gb | BG547965.1 | BG547965 602576071F1 NIH_MGC_77 H...

28 6.4

gi | 13030375 | gb | BG281450.1 | BG281450 602401966F1 NIH_MGC_20 H...

28 6.4

gi | 12951460 | emb | AL582959.1 | AL582959 AL582959 LTI_NFL010_BC2...

28 6.4

gi | 12764352 | gb | BG254536.1 | BG254536 602368464F1 NIH_MGC_91 H...

28 6.4

gi | 12378592 | gb | BF961317.1 | BF961317 PM3-NN0223-111200-004-d0...

28 6.4

gi | 12374538 | gb | BF957263.1 | BF957263 PM3-NN0223-241100-002-b0...

28 6.4

gi | 12323114 | gb | BF926150.1 | BF926150 CM2-NT0193-301100-562-a1...

28 6.4

gi | 12259862 | gb | BF869732.1 | BF869732 IL3-ET0114-251000-316-A1...

28 6.4

gi | 12129894 | gb | BF800905.1 | BF800905 PM1-CI0110-201000-003-f0...

28 6.4

gi | 12071436 | gb | BF744760.1 | BF744760 QV2-BT0635-311000-440-c1...

28 6.4

gi | 11770407 | gb | BE965733.2 | BE965733 601659792R1 NIH_MGC_70 H...

28 6.4

gi | 11766539 | gb | BE963121.2 | BE963121 601656923R1 NIH_MGC_67 H...

28 6.4

gi | 10348536 | gb | BE890328.1 | BE890328 601431783F1 NIH_MGC_72 H...

28 6.4

gi|10142985|gb|BE728993.1|BE728993 601562251F1 NIH_MGC_20 Ho...

28 6.4

gi|10095527|gb|BE707262.1|BE707262 PM1-HT0452-060700-008-e0...

28 6.4

gi|9772196|gb|BE543551.1|BE543551 601070523F1 NIH_MGC_12 Ho...

28 6.4

gi|9768571|gb|BE539926.1|BE539926 601060667F2 NIH_MGC_10 Ho...

28 6.4

gi|9342607|gb|BE397242.1|BE397242 601290754F1 NIH_MGC_8 Ho...

28 6.4

gi|9332870|gb|BE387505.1|BE387505 601274247F1 NIH_MGC_20 Ho...

28 6.4

gi|8140649|gb|AW950985.1|AW950985 EST363055 MAGE resequence...

28 6.4

gi|8139665|gb|AW950129.1|AW950129 EST362094 MAGE resequence...

28 6.4

gi|6879658|gb|AW375004.1|AW375004 MR0-CT0068-280999-002-f07...

28 6.4

gi|5435227|emb|AL079651.1|AL079651 DKFZp434N0629_r1 434 (sy...

28 6.4

gi|5406349|emb|AL036861.1|AL036861 DKFZp56401963_r1 564 (sy...

28 6.4

gi|2566893|gb|AA641675.1|AA641675 nr62g01.s1 NCI_CGAP_Lym3 ...

28 6.4

gi|2080087|gb|AA418268.1|AA418268 zv96d09.s1 Soares_NhHMPu...

28 6.4

gi|2056455|gb|AA402650.1|AA402650 zu49g06.r1 Soares ovary t...

28 6.4

gi|1516398|gb|AA040102.1|AA040102 zk46e02.r1 Soares_pregn...

28 6.4

Alignments

>gi|19365764|gb|BM915385.1|BM915385 AGENCOURT_6701642 NIH MGC 41

Homo sapiens cDNA clone

IMAGE:5481560 5'

Length = 1086

Score = 32.2 bits (16). Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

Subj: 23-acctccctccgcggag 38-

>gi|19353768|gb|BM903897.1|BM903897 AGENCOURT 6696012 NIH MGC 67

Homo sapiens cDNA clone IMAGE:5492392

5'

Length = 1497

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcggag 16

1 1 1 1 1 1 1 1 1 1 1 1 1 1

Subj: 445 acctccctccgcggag 430

>gi|18807810|gb|BM562052.1|BM562052 AGENCOURT 6562015 NIH MGC 118

Homo sapiens cDNA clone

IMAGE:5745414 5'.

Length = 1175

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||

Sbjct: 20 acctccctccgcggag 35

>gi|18791603|gb|BM553137.1|BM553137 AGENCOURT_6572574 NIH_MGC_41

Homo sapiens cDNA clone

IMAGE:5467063 5'.

Length = 1100

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||||||||||||||

Sbjct: 26 acctccctccgcggag 41

>gi|16171065|gb|BI908151.1|BI908151 603067456F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5216508 5'.

Length = 706

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 25 acctccctccgcggag 40

>gi|15759271|gb|BI767693.1|BI767693 603060648F1 NIH_MGC_122 Homo sa
piens cDNA clone TMAGE:5209978 5'.

Length = 862

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 705 acctccctccgcggag 720

>gi|15343643|gb|BI518851.1|BI518851 603061760F1 NIH_MGC_118 Homo sa
piens cDNA clone IMAGE:5210943 5'.

Length = 943

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 25 acctccctccgcggag 40

>gi|14309343|gb|BG899094.1|BG899094 HOA21-1-G9 HOA (Human Osteoarthritic Cartilage) Homo sapiens

cDNA.

Length = 364

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 83 acctccctccgcggag 98

>gi|13662542|gb|BG611171.1|BG611171 602612144F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4737466 5'.

Length = 897

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcggag 16

|||||||||||||||

Sbjct: 809 acctccctccgcggag 794

>gi|12609210|gb|BG115704.1|BG115704 602317174F1 NIH_MGC_88 Homo sa
piens cDNA clone IMAGE:4417482 5'.

Length = 838

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcccggag 16

||||| ||||| |||||

Sbjct: 51 acctccctccgcccggag 66

>gi|12101282|gb|BF796228.1|BF796228 602258513F1 NIH_MGC_85 Homo sa
piens cDNA clone IMAGE:4341962 5'.

Length = 1081

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcccggag 16

||||| ||||| |||||

Sbjct: 7 acctccctccgcccggag 22

>gi|11152079|gb|BF238160.1|BF238160 601811886F1 NIH_MGC_48 Homo sa
piens cDNA clone IMAGE:4054821 5'.

Length = 811

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

|| | | | | | | | | | | |

Sbjct: 11 acctccctccgcggag 26

>gil11100313|gb|BF206727.1|BF206727 601871105F1 NIH_MGC_19 Homo sa
piens cDNA clone IMAGE:4101600 5'.

Length = 888

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||| | | | | | |

Sbjct: 32 acctccctccgcggag 47

>gil11100272|gb|BF206686.1|BF206686 601871051F1 NIII MGC_19 Homo sa
piens cDNA clone IMAGE:4101517 5'.

Length = 917

Score = 32.2 bits (16), Expect = 0.41

Identities = 16/16 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcggag 16

||||| | | | | | |

Sbjct: 33 acctccctccgcccggag 48

>gi|16775383|gb|BM046103.1|BM046103 603625849F1 NIH_MGC_40 Homo s
apiens cDNA clone IMAGE:5452309 5'.

Length = 869

Score = 30.2 bits (15), Expect = 1.6

Identities = 15/15 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcccggag 16

|||||||||||||||

Sbjct: 692 cctccctccgcccggag 706

>gi|19739174|gb|BQ014273.1|BQ014273 UI-H-ED1-axs h-21-0-UI.s1 NCI
CGAP ED1 Homo sapiens cDNA clone

IMAGE:5833028 3'.

Length = 772

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcccggag 15

|||||||||||||||

Sbjct: 495 cctccctccgcccggag 482

>gi|19378603|gb|BM928224.1|BM928224 AGENCOURT 6699855 NIH_MGC_121
Homo sapiens cDNA clone IMAGE:5770072 5'.

Length = 1140

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||| : |||||

Sbjct: 1009 cctccctccgcgga 1022

>gi|19367808|gb|BM917429.1|BM917429 AGENCOURT_6606724 NIH_MGC_106

Homo sapiens cDNA clone IMAGE:5483947 5'.

Length = 1073

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||| : |||||

Sbjct: 916 acctccctccgcgg 929

>gi|19364214|gb|BM913835.1|BM913835 AGENCOURT_6612786 NIH_MGC_98

Homo sapiens cDNA clone IMAGE:5477539 5'.

Length = 1104

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcgga 15

|||||||||||||

Sbjct: 842 cctccctccgcgga 829

>gi|19361343|gb|BM910964.1|BM910964 AGENCOURT_6615957 NIH_MGC_98

Homo sapiens cDNA clone IMAGE:5454547 5'.

Length = 1128

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 3 ctccctccgcggag 16

|||||||T|||||||

Sbjct: 883 ctccctccgcggag 870

>gi|18505954|gb|BM456914.1|BM456914 AGENCOURT_6404253 NIH_MGC_92

Homo sapiens cDNA clone

IMAGE:5583862 5'.

Length = 1813

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

|||||||T|||||||

Sbjct: 29 cctccctccgcgga 42

>gi|18499709|gb|BM450669.1|BM450669 AGENCOURT_6394717 NIH_MGC_67

Homo sapiens cDNA clone IMAGE:5494366 5'.

Length = 1430

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 1150 acctccctccgcgg 1163

>gi|16000196|gb|BI859449.1|BI859449 603388188F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5396997 5'.

Length = 852

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 100 acctccctccgcgg 113

>gi|15928460|gb|BI818193.1|BI818193 603032663F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5173838 5'.

Length = 683

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcca 15

||||||| | | | | |

Sbjct: 96 cctccctccgcca 109

>gi|15431547|gb|BI544235.1|BI544235 603241605F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5284296 5'.

Length = 676

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 3 ctccctccgcca 16

||||||| | | | | |

Sbjct: 39 ctccctccgcca 26

>gi|15345229|gb|BI520437.1|BI520437 603071622F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5163773 5'.

Length = 727

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcca 14

||||||||||||||

Sbjct: 505 acctccctccgcccgg 492

>gi|14440373|gb|BI033747.1|BI033747 PM3-NN0223-220201-014-h04 NN022

3 Homo sapiens cDNA.

Length = 284

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcccgg 14

||||||||||||||

Sbjct: 97 acctccctccgcccgg 84

>gi|14426676|gb|BI020046.1|BI020046 CM3-MT0291-110101-622-f04 MT029

1 Homo sapiens cDNA.

Length = 436

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcccgg 14

||||||||||||||

Sbjct: 365 acctccctccgcccgg 352

>gi|14081325|gb|BG770672.1|BG770672 602734012F1 NIH_MGC_49 Homo sa

piens cDNA clone IMAGE:4859546 5'.

Length = 949

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

|||||| '|||||

Sbjct: 63 acctccctccgcgg 76

>gi|13546630|gb|BG547965.1|BG547965 602576071F1 NIH_MGC_77 Homo sa
piens cDNA clone IMAGE:4704209 5'.

Length = 918

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||| | | | | |

Sbjct: 248 acctccctccgcgg 261

>gi|13030375|gb|BG281450.1|BG281450 602401966F1 NIH_MGC_20 Homo sa
piens cDNA clone IMAGE:4544201 5'.

Length = 782

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

|||||||||||||

Sbjct: 417 acctccctccgcgg 430

>gi|12951460|emb|AL582959.1|AL582959 AL582959 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL008YA12 3' prime.

Length = 822

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcgga 15

|||||||||||||

Sbjct: 533 cctccctccgcgga 520

>gi|12764352|gb|BG254536.1|BG254536 602368464F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4476902 5'.

Length = 1031

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

|||||||||||||

Sbjct: 849 cctccctccgcgga 862

>gi|12378592|gb|BF961317.1|BF961317 PM3-NN0223-111200-004-d03 NN0

223 Homo sapiens cDNA.

Length = 277

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 89 acctccctccgcgg 76

>gi|12374538|gb|BF957263.1|BF957263 PM3-NN0223-241100-002-b08 NN0

223 Homo sapiens cDNA.

Length = 168

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 117 acctccctccgcgg 104

>gi|12323114|gb|BF926150.1|BF926150 CM2-NT0193-301100-562-a12 NT01

93 Homo sapiens cDNA.

Length = 417

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 268 cctccctccgcgga 255

>gi|12259862|gb|BF869732.1|BF869732 IL3-ET0114-251000-316-A11 ET01
14 Homo sapiens cDNA.

Length = 278

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgccc 14

||||||||||||||

Sbjct: 73 acctccctccgccc 60

>gi|12129894|gb|BF800905.1|BF800905 PM1-CI0110-201000-003-f08 CI01
10 Homo sapiens cDNA.

Length = 283

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Quqry: 1 acctccctccgccc 14

|||||||||| |||

Sbjct: 211 acctccctccgcgg 224

>gi|12071436|gb|BF744760.1|BF744760 QV2-BT0635-311000-440-c11 BT06
35 Homo sapiens cDNA.

Length = 534

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

|||||||

Sbjct: 319 cctccctccgcgga 332

>gi|11770407|gb|BE965733.2|BE965733 601659792R1 NIH_MGC_70 Homo sa
piens cDNA clone IMAGE:3896134 3'.

Length = 1336

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||| |

Sbjct: 292 acctccctccgcgg 305

>gi|11766539|gb|BE963121.2|BE963121 601656923R1 NIH_MGC_67 Homo sa
piens cDNA clone IMAGE:3865924 3'.

Length = 1442

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 3 ctccctccgcggag 16

||||||| | | | | |

Sbjct: 403 ctccctccgcggag 390

>gi|10348536|gb|BE890328.1|BE890328 601431783F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3916820 5'.

Length = 794

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||| | | | | |

Sbjct: 115 acctccctccgcgg 128

>gi|10142985|gb|BE728993.1|BE728993 601562251F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3831924 5'.

Length = 840

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||| |||||

Sbjct: 397 acctccctccgcgg 410

>gi|10095527|gb|BE707262.1|BE707262 PM1-HT0452-060700-008-e08 HT04

52 Homo sapiens cDNA.

Length = 592

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Minus

Query: 1 acctccctccgcgg 14

||||||| |||||

Sbjct: 343 acctccctccgcgg 330

>gi|9772196|gb|BE543551.1|BE543551 601070523F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456940 5'.

Length = 1035

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||| |||||

Sbjct: 332 acctccctccgcgg 345

>gi|9768571|gb|BE539926.1|BE539926 601060667F2 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3456940 5'.

iens cDNA clone IMAGE:3447161 5'.

Length = 902

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 411 acctccctccgcgg 424

>gi|9342607|gb|BE397242.1|BE397242 601290754F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3621253 5'.

Length = 524

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 228 cctccctccgcgga 241

>gi|9332870|gb|BE387505.1|BE387505 601274247F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3615538 5'.

Length = 637

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 1 acctccctccgcgg 14

||||||||||||||

Sbjct: 422 acctccctccgcgg 435

>gi|8140649|gb|AW950985.1|AW950985 EST363055 MAGE resequences, MA

GA Homo sapiens cDNA.

Length = 638

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

||||||||||||||

Sbjct: 273 cctccctccgcgga 286

>gi|8139665|gb|AW950129.1|AW950129 EST362094 MAGE resequences, MAG

A Homo sapiens cDNA.

Length = 611

Score = 28.2 bits (14), Expect = 6.4

Identities = 14/14 (100%)

Strand = Plus / Plus

Query: 2 cctccctccgcgga 15

|||||||:|||||

Sbjct: 273 cctccctccgcgga 286

Database: GenBank Human EST entries

Posted date: Mar 29, 2002 2:35 AM

Number of letters in database: 2,114,234,064

Number of sequences in database: 4,280,058

Lambda K H

1.37 0.711 1.31

Gapped

Lambda K H

1.37 0.711 1.31

Matrix: blastn matrix:1 -3

Gap Penalties: Existence: 5, Extension: 2

Number of Hits to DB: 5013

Number of Sequences: 4280058

Number of extensions: 5013

Number of successful extensions: 5013

Number of sequences better than 10.0: 61

length of query: 16

length of database: 2,114,234,064

effective HSP length: 15

effective length of query: 1

effective length of database: 2,050,033,194

effective search space: 2050033194

effective search space used: 2050033194

T: 0

A: 30

X1: 6 (11.9 bits)

X2: 15 (29.7 bits)

S1: 12 (24.3 bits)

S2: 14 (28.2 bits)

Top of Form

	<u>Taxono</u>
1: BM915385. AGENCOURT_6701642... [gi:19365764]	<u>my_Lin</u>
	<u>kOut</u>

IDENTIFIERS

dbEST Id: 11598757
 EST name: AGENCOURT_6701642
 GenBank Acc: BM915385
 GenBank gi: 19365764

CLONE INFO

Clone Id: IMAGE:5481560 (5')
 Plate: LLCM2006 Row: d Column: 09
 DNA type: cDNA

PRIMERS

PolyA Tail: Unknown

SEQUENCE

CGCCCTGGGCCATCTCCCTCCCACCTCCCTCCGGAGCAGCCAGACAGCGAGGGC

CCCG

GCCGGGGGCAGGGGGACGCCCGTCCGGGCACCCCCCGGCTTGAGCCGCCCC

CGGG

GCCGGCCTCGGCCGGAGCGGAGGAAGGAGTCGCCAGGAGCAGCTGAGGCCCA

CAGT

CTGAGACGAGCCGCCGCCGCCCCGCCACTGCCGGAGGAGGGGGAGGAGGAGCGG

GAGG

AGGGACGAGCTGGTCGGAGAAGAGGAAAAAACTTTGAGACTTTCCGTTGCCG
CTGG
GAGCCGGAGGCGCGGGGACCTCTTGGCGCGACGCTGCCCCGCGAGGAGGCAGGACT
TGGG
GACCCCAGACCGCCTCCCTTGCCGCCGGGACGCTTGCTCCCTCCCTGCCCTA
CACG
GCGTCCTCAGGCGCCCCATTCCGGACCAGCCCTCGGGAGTCGCCGACCCGGCCT
CTCG
CAAAGACTTTACCATACCTCGGGCGCACCTCTGCACGCCGCTTCATCACCGG
CCTG
TCTACTGAGCCCCCGCGGATGCCTAGACCCCTTCCTCCGGAGACGGATCCCTC
TCCG
ACCTGCCGCAAATTCCCTATTCTGGAACACCCCCGCTTCCTGGGACCTAATCCCC
GCCT
TTCGACGCTCCTTGCCTGGGAACTGAAGAGCCCCGGGTCGTAACCTTTCC
TCCC
CGTTTGAAAAACATCCCCGTTAATAAACCTTACTATTTCGCTTGGCCCC
CCCT
TACGGTTTTGGCGGGCACTAAACAAACATCGAGTCTCAAGGCCGGATGCCACT
CAAG
CCTGAATACTTTGCCGTTAGGGCGGTCTTACCGAGTAGAGTCGGCCTTG
ACCG
GACCCTATTCAATTGGTTCCCGTGACGTGTGCCGCTAACGAGATATTAACCTCT
CCCG
ACACATTGTCATAAAACACCACTTTGACACGCCACTCCTGTTAATAGTCGCC
CCTC
CCCGCGTGTAAAATTCCCGCCAATGCCCTCATTATTCCGCTCCATGAAAAAG
GGGG
TCGGCN

Quality: High quality sequence stops at base: 467

Entry Created: Mar 11 2002

Last Updated: Mar 12 2002

COMMENTS

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I. M. A. G. E. Consortium (LLNL

)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information c

an

be found through the I. M. A. G. E. Consortium/LLNL at:

<http://image.llnl.gov>

LIBRARY

Lib Name: NIH_MGC_41

Organism: Homo sapiens

Organ: skin

Tissue type: amelanotic melanoma, cell line

Lab host: DH10B (phage-resistant)

Vector: pOTB7

R. Site 1: XhoI

R. Site 2: EcoRI

Description: cDNA made by oligo-dT priming. Directionally cloned into
EcoRI/XhoI sites using the following 5' adaptor: GGCACGA

G (G

). Library constructed by Ling Hong in the laboratory of

Gerald M. Rubin (University of California, Berkeley) using

ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT

(Life Technologies). Note: this is a NIH_MGC Library.

SUBMITTER

Name: Robert Strausberg, Ph. D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Institutes of Health, Mammalian Gene Collection (MGC)
Authors: NIH-MGC <http://mgc.nci.nih.gov/>
Year: 1999
Status: Unpublished

Bottom of Form

Revised: October 24, 2001.

Check on Est in Genbank:

Query= (1086 letters)

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, CSS, or phase 0, 1 or 2 HTGS sequences)

1,205,903 sequences; 5,297,758,116 total letters

Score E

Sequences producing significant alignments: (bits)

Value

gi|10863872|ref|NM_000660.1| Homo sapiens transforming grow... 58
7 e-165

gi|18590091|ref|XM_085882.1| Homo sapiens similar to transf... 587
e-165

gi|11424057|ref|XM_008912.1| Homo sapiens transforming grow... 587
e-165

gi|7684381|gb|AC011462.4|AC011462 Homo sapiens chromosome 1... 5
87 e-165

gi|37097|emb|X05839.1|HSTGFBG1 Human transforming growth fa...
587 e-165

gi|37092|emb|X02812.1|HSTGFB1 Human mRNA for transforming g...
587 e-165

gi|340526|gb|J04431.1|HUMTGFB1PR Homo sapiens transforming ...
587 e-165

gi|12654682|gb|BC001180.1|BC001180 Homo sapiens, Similar to... 2
91 8e-76

gi|12652748|gb|BC000125.1|BC000125 Homo sapiens, Similar to... 2
91 8e-76

gi|18490115|gb|BC022242.1| Homo sapiens, clone MGC:22008 IM... 15
3 4e-34

gi|755044|gb|M23703.1|PIGTGFB1A Sus scrofa transforming gro...
129 6e-27

gi|7650477|gb|AF249327.1|AF249327 Rattus norvegicus TGF-bet...
66 8e-08

gi|4416081|gb|AF105069.1|AF105069 Rattus norvegicus transfo...
66 8e-08

gi|2394170|gb|AF015683.1|AF015683 Rattus norvegicus transfo...
66 8e-08

gi|6755774|ref|NM_011577.1| Mus musculus transforming growth factor beta 1... 64

3e-07

gi|1161133|gb|L42456.1|MUSTGFE1G01 Mus musculus TGF-1 gene, ...

64 3e-07

gi|3688423|emb|AJ009862.1|MMU009862 Mus musculus mRNA for transforming growth factor beta 1...

64 3e-07

gi|201947|gb|M57902.1|MUSTGFB1 Mouse transforming growth factor beta 1...

64 3e-07

gi|18042365|gb|AC097483.3| Homo sapiens BAC clone RP11-146N... 4

4 0.30

gi|17481821|ref|XM_008785.3| Homo sapiens one cut domain, family... 44

0.30

gi|12737997|ref|XM_007116.2| Homo sapiens Zic family member... 44

0.30

gi|6005961|ref|NM_007129.1| Homo sapiens Zic family member... 44

0.30

gi|11065969|gb|AF193855.1|AF193855 Homo sapiens zinc finger... 44

44 0.30

gi|4758847|ref|NM_004852.1| Homo sapiens one cut domain, family... 44

0.30

gi|15787728|emb|AL355338.33|AL355338 Human DNA sequence from ... 44

44 0.30

gi|4028591|gb|AF104902.1|AF104902 Homo sapiens ZIC2 protein... 44

44 0.30

gi|1531593|gb|U50523.1|HSU50523 Human BRCA2 region, mRNA se... 44

44 0.30

gi|4468940|emb|Y18198.1|HSAY18198 Homo sapiens mRNA for ONE... 44

44 0.30

gi|19067958|gb|AY049805.1| Alopia pelagicus 5.8S ribosomal... 42

1. 2

gi|18025465|gb|AY037858.1| Cercopithicine herpesvirus 15 st... 42

1. 2

gi|12039248|gb|AC020659.5|AC020659 Homo sapiens chromosome ... -

42 1. 2

gi|19909461|gb|AC098709.3| Mus musculus clone RP23-1K14, co... 4

0 4. 6

gi|19921137|ref|NM_135651.1| Drosophila melanogaster (CG47... 4

0 4. 6

gi|18376846|gb|AC092198.2| Homo sapiens chromosome X clone ... 4

0 4. 6

gi|18467841|ref|XM_078995.1| CG4751 (CG4751), mRNA 40

4. 6

gi|18376869|gb|AC091898.2| Homo sapiens chromosome 5 clone ... 4

0 4. 6

gi|18030132|gb|AC026695.5| Homo sapiens chromosome 5 clone ... 4

0 4. 6

gi|15887302|gb|AC020914.8| Homo sapiens chromosome 19 clone... 4

0 4. 6

gi|14578122|gb|AC092241.1|AC092241 Drosophila melanogaster,...

40 4. 6

gi|15292266|gb|AY051978.1| Drosophila melanogaster LD44770 ... 40

4. 6

gi|15055218|gb|AC060226.39| Homo sapiens 12 BAC RP11-101P14... 4

0 4. 6

gi|14389338|gb|AC084282.6|AC084282 Oryza sativa chromosome ... -

40 4. 6

gi|13677167|gb|AC015977.9|AC015977 Homo sapiens clone RP11-...

40 4. 6

gi|9910225|ref|NM_020179.1|| Homo sapiens FN5 protein (FN5),... 40
4.6

gi|10440613|gb|AC069145.5|AC069145 Oryza sativa chromosome ...
40 4.6

gi|10728714|gb|AE003631.2|AE003631 Drosophila melanogaster ...
40 4.6

gi|9246422|gb|LAF197137.1|LAF197137 Homo sapiens FN5 protein ...
40 4.6

gi|4190938|gb|AC000091.1|AC000091 Homo sapiens Chromosome 2...
40 4.6

gi|17431932|emb|AL646085.1|AL646085 Ralstonia solanacearum ...
40 4.6

gi|15073719|emb|AL591785.1|SME591785 Sinorhizobium meliloti...
40 4.6

gi|3628578|gb|AC005115.1|AC005115 Drosophila melanogaster D...
40 4.6

gi|3150432|gb|U50080.1|LSU50080 Lymnaea stagnalis serotonin...
40 4.6

gi|8052359|emb|AL356592.1|SC9H11 Streptomyces coelicolor co...
40 4.6

gi|6624640|cmb|AL084344.24|HS118B18 Human DNA sequence from...
40 4.6

gi|15528721|dbj|AP003296.3| Oryza sativa (japonica cultivar... 40
4.6

gi|15289781|dbj|AP003141.2| Oryza sativa (japonica cultivar... 40
4.6

gi|6069643|dbj|AP000616.1| Oryza sativa (japonica cultivar... 40
4.6

gi|960285|gb|L46862.1|RATLAMB2G Rattus norvegicus laminin B...
4.6

40 4.6

gi|198704|gb|J03749.1|MUSLAMB2B Mouse laminin B2 gene, exon...

40 4.6

gi|198702|gb|J02930.1|MUSLAMB2A Mouse laminin B2 chain mRNA...

40 4.6

gi|198694|gb|J03484.1|MUSLAM2B Mouse laminin B2 chain mRNA,...

40 4.6

Alignments

>gi|10863872|ref|NM_000660.1| Homo sapiens transforming growth factor, beta 1 (Camurati-Engelmann

disease) (TGFB1), mRNA

Length = 2745

Score = 587 bits (296), Expect = e-165

Identities = 356/377 (94%), Gaps = 1/377 (0%)

Strand = Plus / Plus

Query: 246 cgagctggtcggagaagagggnnnnnnctttgagactttccgttgcgtggagcc
305

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Sbjct: 225 cgagctggtcggagaagaggaaaaactttgagactttccgttgcgtggagcc
284

Query: 306 ggaggcgcggggacctcttggcgacgctgccccgaggaggcaggacttggggaccc
365

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Sbjct: 285 ggaggcgcggggacctcttggcgacgctgccccgaggaggcaggacttggggaccc
344

Query: 366 cagaccgcctcccttgcgcggggacgcttgcctccctgcacacggcgtc
425
||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct: 345 cagaccgcctcccttgcgcggggacgcttgcctccctgcacacggcgtc
404
||||||||||||||||||||||||||||||||||||||||||||||||||||

Query: 426 cctcaggcgcccccattccggaccagccctcgaggatcgccgaccggcctcgcaaag
485
||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct: 405 cctcaggcgcccccattccggaccagccctcgaggatcgccgaccggcctccgcaaag
464
||||||||||||||||||||||||||||||||||||||||||||||||

Query: 486 acttttaccataacctcgccgcaccctctgcacgcggcattcatcaccggctgtctac
545
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct: 465 actttttcccaagacctcgccgcacccctgcacgcggcattcatccggcctgtctcc
524
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query: 546 tgagcccccgcggatgcctagacccttctcctccggagacggatccctctccgacctg
605
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct: 525 tgagcccccgcgcattccttagacccttctcctccaggagacggatctctccgacctg
583
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query: 606 ccgcaaattccctattc 622
||| ||| ||| |||

Sbjct: 584 ccacagatccctattc 600
【0083】

実施例6： 5'末端配列タグの統計解析

1つの試料中の同一の複数のmRNAから得られた5'末端配列タグ又は同一のcDNAライブラリー中の核酸断片は、実施例5に記載したように、NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) のような標準的なソフトウェアソリューションによって分析し、唯一の配列タグの同定を行うことができる。このような唯一の配列タグの全てを次いで個々に計数し、同一の試料からえらレア全てのタグの総数に治する個々の唯一のタグの寄与を分析することができる。全てのタグの総数に対する個々のタグの寄与は、その試料又はcDNAライブラリー中の複数のmRNAの転写物の定量を可能にする。このようにして得られた、個々の試料についての結果は、他の試料から得られた同様なデータと比較することにより、それらの発現パターンを比較することができる。

【0084】

実施例7： 5'末端配列タグのゲノム配列情報へのマッピング

この実施例で記載したようにして得られた5'末端特異配列タグは、部分又は全配列が得られているゲノム中の転写された領域を同定するために用いることができる。このような検索は、NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) のような標準的なソフトウェアソリューションを用いて、5'末端特異配列タグをゲノム配列に対して整列させることにより行うことができる。ヒト、ラット、マウスのような、巨大なゲノムの場合には、例えば実施例#に記載されたアプローチによって連結体から得られた最初の配列情報を伸長することが必要になるかもしれない。伸長された配列を用いることにより、ゲノム中の活性に転写される領域のより精密な同定が可能になる。

【0085】

実施例8： 停車開始部位の同定

ゲノム配列に対してマッピングすることができる5'末端特異配列タグは、調節配列の同定を可能にする。遺伝子において、転写流域の5'末端よりも上流のDNAは、通常、遺伝子発現をコントロールするために用いられる、ほとんどの調節要素を含む。これらの調節配列は、さらに、転写因子の結合部位の情報を有

するデータベースの検索により、さらに分析することができる。プロモーター解析のための、転写因子結合部位についての公衆が利用可能なデータベースとして次のものを挙げることができる。

Transcription Regulatory Region Database (TRRD)

(<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trrd4/>)

TRANSFAC (<http://transfac.gbf.de/TRANSFAC/>)

TFSEARCH (<http://www.cbrc.jp/research/db/TFSEARCH.html>)

PromoterInspector provide by Genomatix Software

(<http://www.genomatix.de/>)

【0086】

実施例9： 5'末端配列タグから得られる情報を用いた全長cDNAのクローニング

連結体から得られる配列情報は、全長cDNAのクローニングのための特異的プライマーの合成に用いることができる。このようなアプローチでは、ある5'末端特異タグから誘導された配列は、フォワードプライマーの設計に用いることができ、一方、リバース側プライマーの選択は、増幅反応に用いられる鑄型DNAに依存する。ポリメラーゼ連鎖反応（PCR）による増幅を、生物学的試料及びオリゴdTプライマーから得られた複数のRNAから誘導された鑄型を用いて行うことができる。第1工程では、オリゴdTプライマーと逆転写酵素を用いてcDNAプールを合成する。第2工程で、5'末端特異タグから誘導されたフォワードプライマーと、オリゴdTプライマーとを用いて、cDNAプールからの全長cDNAの増幅に用いる。同様に、5'末端タグから誘導されたフォワードプライマーと、ベクター内に位置するリバースプライマーとを用いて、既存のcDNAライブラリーから特定の全長cDNAを増幅することができる。

【0087】

実施例10： cDNAライブラリーからの5'末端タグのクローニングの代替アプローチ

挿入物の5'末端にクラスIISエンドヌクレアーゼの認識部位を有する、既存のcDNAライブラリーから、複数のcDNAを増幅することができる。この

ようなライブラリーから誘導されたP C R産物は、ここに記載した実施例のようにさらに処理することができる。

【0088】

実施例11： クラスI I s認識部位を有するオリゴヌクレオチドによるキャップ構造の交換による5'末端のクローニング

最初の転写物の5'末端を含むc DNA/RNAハイブリッドは、実施例1に記載したようにして得ることができる。このようなc DNA/RNAハイブリッド中のキャップ構造は、次に、T 4 ポリヌクレオチドキナーゼ又はタバコアシッドピロフォスファターゼのような加水分解酵素により、酵素的に除去することができる。クラスI I s認識部位を有する一本鎖又は二本鎖オリゴヌクレオチドを次にT 4 RNAリガーゼにより、キャップを除去したm RNAの5'末端に位置するリン酸に連結する。連結されたオリゴヌクレオチドは、実施例1に記載した手順に従い、第2鎖合成のプライマーとして機能するであろう。ライゲーション工程において修飾オリゴヌクレオチドを用いることにより、二本鎖c DNAを支持体に結合することができ、ここに記載したように連結体のクローニングに用いることができる。

【0089】

実施例12： 試料の増幅工程

試料の量が本発明にとって限定されている場合には、試料材料を次のアプローチにより増幅することができる。第1工程で、実施例11に記載したように複数のm RNAを処理することにより、キャップ構造を、クラスI I s認識部位を有する適当なオリゴヌクレオチドに交換する。第2工程で、リンカーに相補的なプライマーと、ポリAプライマーとを用いたP C Rにより、上記鑄型を増幅する。P C R産物は、実施例1に記載したように本発明に用いることができる。

【0090】

実施例13： 伸長された5'末端配列の利用

連結体について得られた、最初の5'末端配列を用いてシーケンシングプライマーを合成し、転写領域の5'末端の伸長された配列情報を得ることができる。

【0091】

実施例14： 遺伝子不活性化

5' 末端特異配列タグから得られた配列情報は、アンチセンスプローブの設計に用いることができ、これは次にノックダウン研究に適用することができる。

【0092】

【発明の効果】

本発明により、試料中に含まれるmRNAの塩基配列情報を取得できるのみならず、新規遺伝子のクローニングを行うことも可能にする新たな手段が提供された。本発明の方法によれば、試料中の複数のmRNAの5'末端領域の塩基配列情報を効率的に得ることができる。5'末端領域の塩基配列情報を得ることができるので、未知の遺伝子のクローニングも行うことができる。さらには、転写開始部位のマッピング、プロモーター利用（ユーセージ）パターンのマッピング、プロモーターのSNPsの解析、発現解析、オルタナティブプロモーターユーセージ及び他のデータとの組合せによる遺伝子ネットワークの創製、断片化ゲノミックDNA中のプロモーター領域の選択的回収等も可能になる。

【0093】

特に、本発明は、プロモーター領域の同定、クローニング及び更なる解析い大きなインパクトを有する。複数の5'末端の情報を有する連結体ライブラリーをシーケンシングすると、転写開始部位の分布を統計学的に解析することが可能になるであろう。異なる生理学的条件間の変化がmRNA転写機械を新しい「状態」にスイッチする。このような「転写状態」は、（1）転写開始部位の存在、（2）タグを計数し、開始部位の存在及び転写因子を関連付けることによりそれらの発現を計数することにより種々の転写因子のデジタル発現をコンピューティングすることにより測定することができる。2つの異なる条件間の遺伝子発現の同様を比較することにより、遺伝子ネットワーキングに関する更なる情報が得られるであろう。種々の疾病及び正常組織間の転写条件のこのような比較は、新規な、極めて理解しやすい診断ツールの設計を可能にする。したがって、本発明は、遺伝子の発見及び解析における高い商業価値を有し、本発明は、新規な診断及び治療産物の開発に用いられるであろう。

【0094】

[文献]

[REFERENCES]

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW, Serial analysis of gene expression, Science 1995 Oct 20;270(5235):484-7

US patent 5,866,330 (SAGE)

US patent 5,695,937 (SAGE)

Piero CARNINCI et al., METHODS IN ENZYMOLOGY, VOL. 303, pp. 19-44, 1999

Lee S, Clark T, Chen J, Zhou G, Scott LR, Rowley JD, Wang SM, Correct identification of genes from serial analysis of gene expression tag sequences, Genomics 2002 Apr;79(4):598-602

Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE, Using the transcriptome to annotate the genome, Nat Biotechnol 2002 May;20(5):508-12

Maruyama K and Sugano S, Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene. 1994, Vol. 138:171-4

Edery I, Chu LL, Sonenberg N, Pelletier J, An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture), Mol Cell Biol 1995 Jun;15(6):3363-71

US patent 6,022,715 (GenSet)

Shibata Y, Carninci P, Watahiki A, Shiraki T, Konno H, Muramatsu M, Hayashizaki Y, Cloning full-length, cap-trapper-selected cDNAs by using the single-strand linker ligation method, Biotechniques 2001 Jun;30(6):1250-4

Sambrook J and Russel DW, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, 2001

Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watahiki A, Shibata K, Konno H, Muramatsu M, Hayashizaki Y, Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis, Genomics. 2001 Sep;77(1-2):79-90.

Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatjeva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, Podkolodny NL, Kolchanov NA, Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL, Nucleic Acids Res 1998 Jan 1;26(1):362-7

Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene. 1994 Jan 28;138(1-2):171-4.

Jordan B., DNA Microarrays: Gene Expression Applications, Springer-Verlag, Berlin Heidelberg New York, 2001

Schena A, DNA Microarrays, A Practical Approach, Oxford University Press, Oxford 1999

US patent 5,962,272 (Clontech)

Carninci P, Shiraki T, Mizuno Y, Muramatsu M, Hayashizaki Y, Extra-long first-strand cDNA synthesis, Biotechniques 2002 May;32(5):984-5

US patents 6,352,828; 6,306,597; 6,280,935; 6,265,163; 5,695,934 (Lynx)

【図面の簡単な説明】

【図 1】

複数の第一鎖の作成の一例が示されている。出発材料は、生物学的試料から得られるRNA又はcDNAライブラリーであり得る。

【図 2】

5'末端特異タグを連結体にクローニングする一例が示されている。この例は、制限酵素Gsu I, Bgl I I 及びEco RIの使用を含むが、これらに限定されるものではない。

【図 3】

第1のリンカーを、5'末端特異タグのクローニングのために用いる例が示されている。この例は、制限酵素Gsu I, Bgl I I 及びEco RIの使用を含むが、これらに限定されるものではない。

【図 4】

第2のリンカーを、5'末端特異ジタグのクローニングのために用いる例が示

されている。この例は、制限酵素G s u I, B g 1 I I 及びE c o R I の使用を含むが、これらに限定されるものではない。

【図5】

ジタグの構造の1例が示されている。この例は、制限酵素G s u I, B g 1 I I 及びE c o R I の使用を含むが、これらに限定されるものではない。

【図6】

5' 末端特異リンカーの用途の一例が示されている。この例では、リンカーは、個々の核酸の豊富化及びそれらのシーケンシングのために用いられる。

【書類名】 図面

【図1】

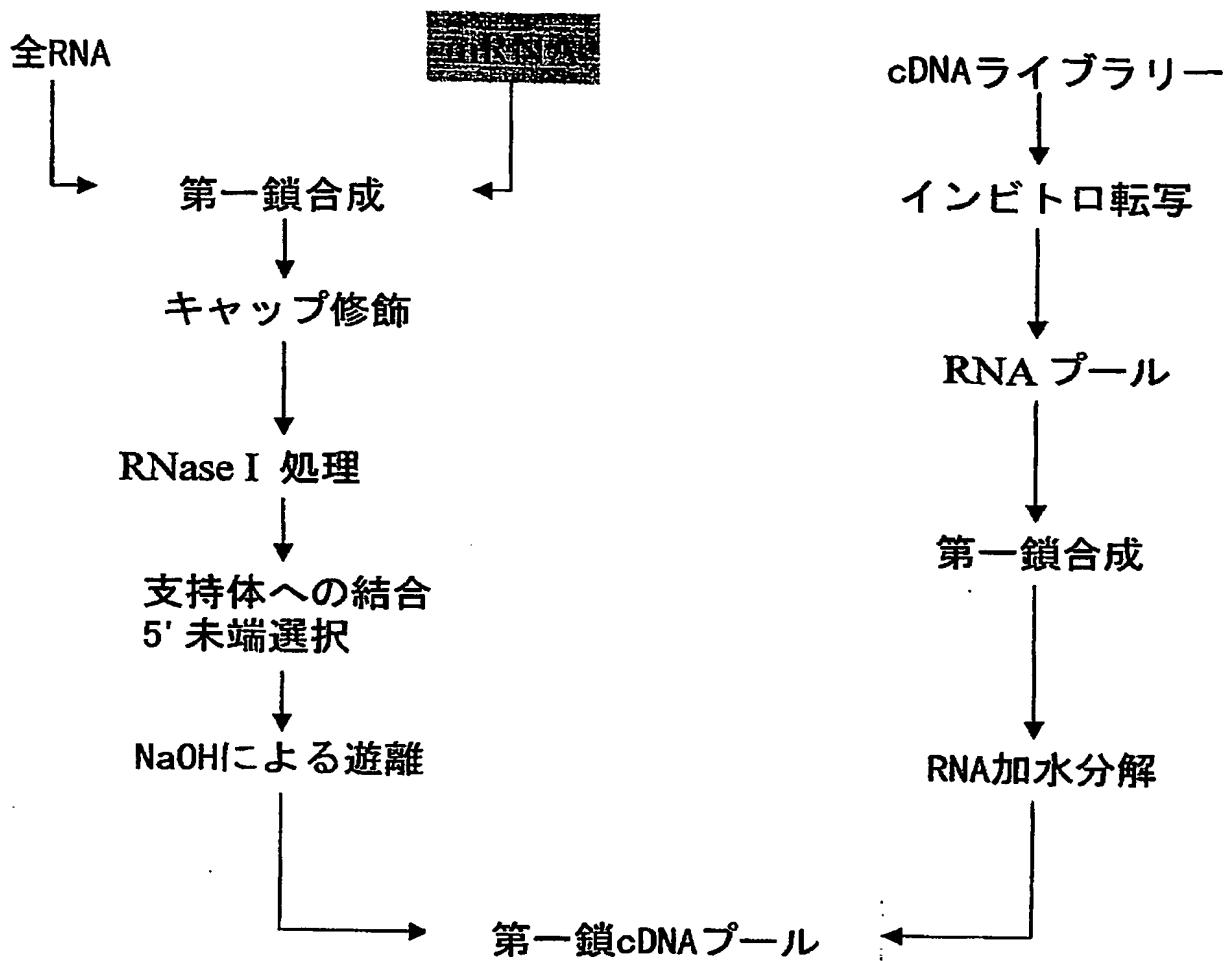


図1:第一鎖合成

【図2】

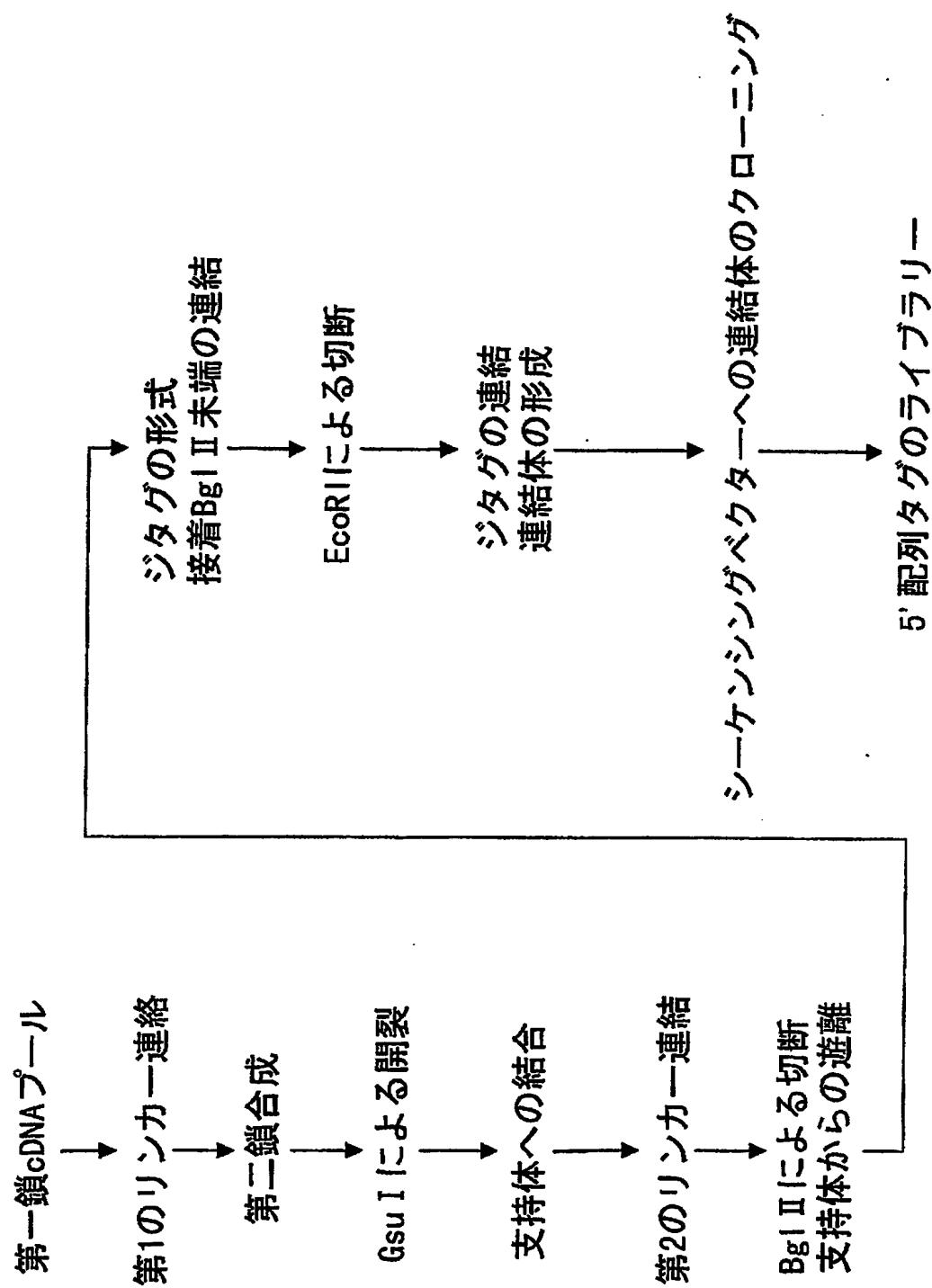
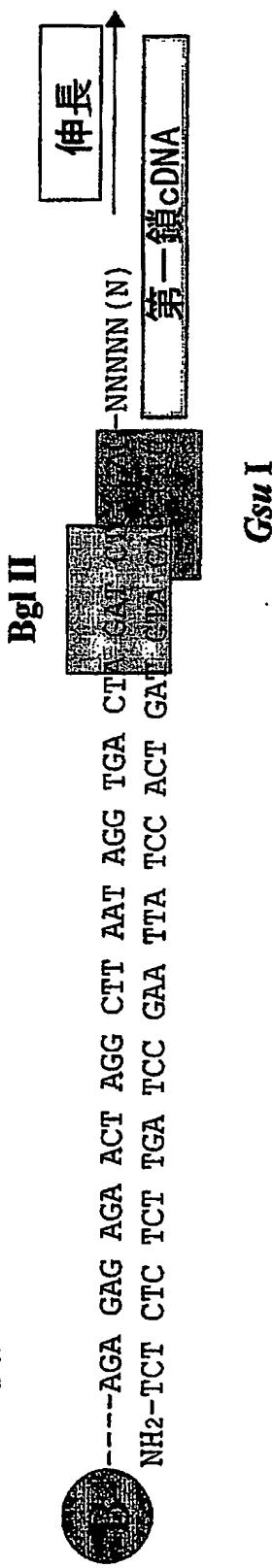


図2:連結体のクローニング

【図3】

Gsu I の使用のために



Gsu I の使用のために

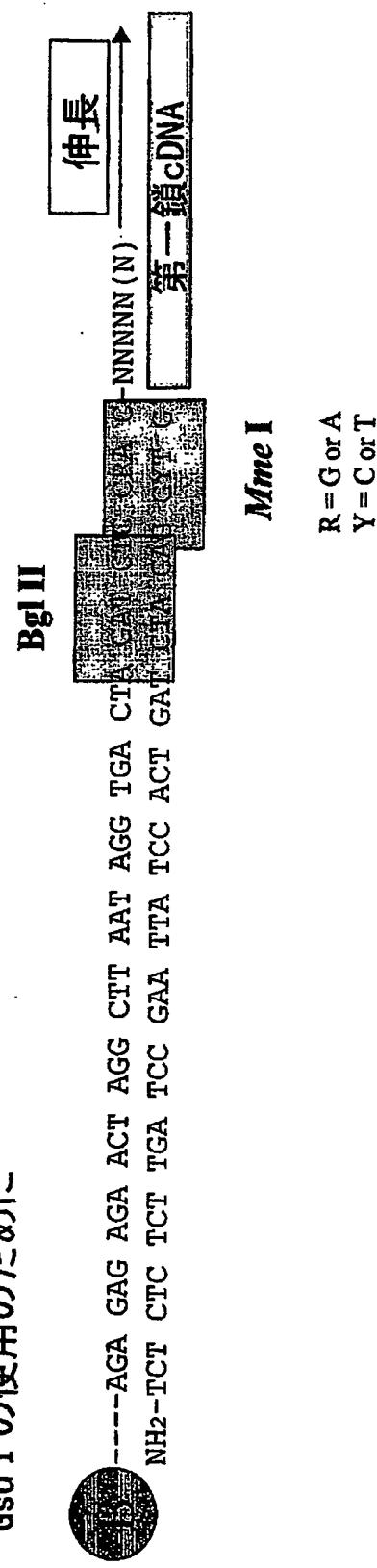


図3: 第一のリシンカ一連結

【図4】

Gsu I の使用のために

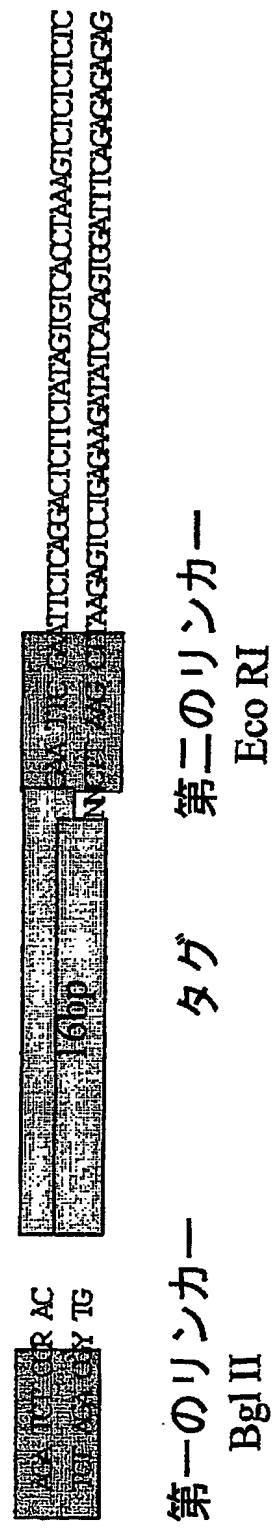
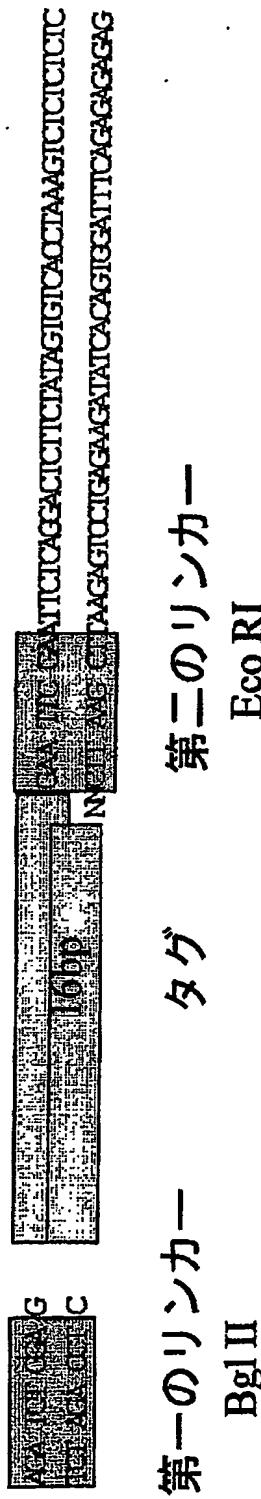
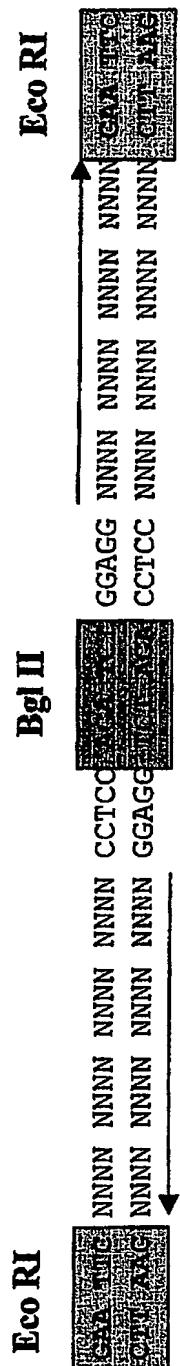


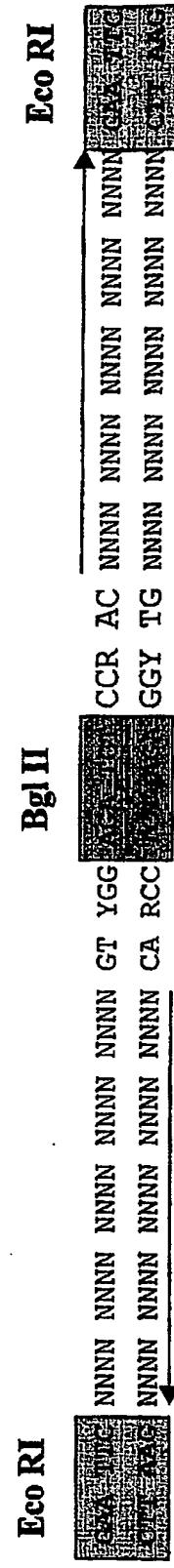
図4: 第二のリンクマーー連結

【図5】

Gsu I の使用のためには



Mme I の使用のためには



R = G or A
Y = C or T

図5. ジタグの構造

【図6】

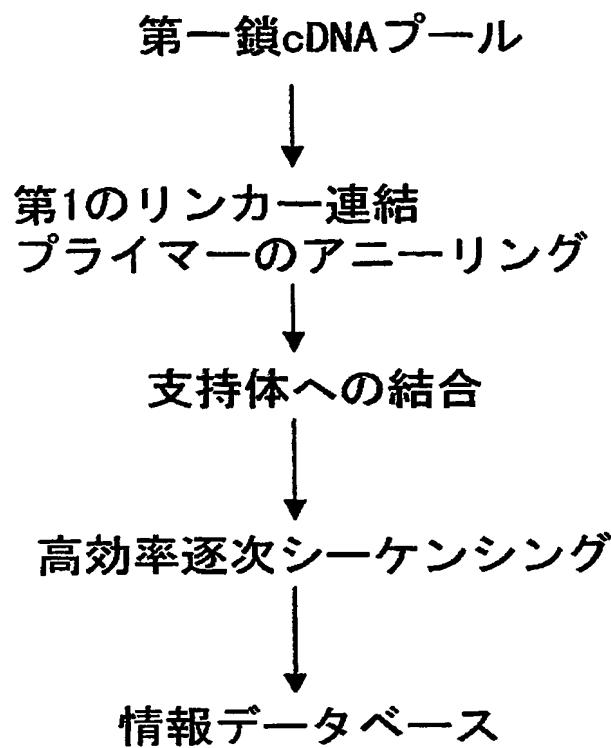


図6:逐次シーケンシング

【書類名】 要約書

【要約】 生物学的材料又は合成プールから得られた複数の核酸断片からの転写領域の5'末端を得る方法が開示されている。5末端をコードするDNA断片が、それらの個々の分析のために、又はそれらの連結体の解析のために豊富化される。5末端から得られた配列情報は、転写物の特徴づけとクローニングのため用いることができる。

【選択図】 なし

認定・付加情報

特許出願の番号 特願2002-235294
受付番号 20201970126
書類名 翻訳文提出書
担当官 田丸 三喜男 9079
作成日 平成15年 2月24日

<認定情報・付加情報>

【提出日】 平成14年10月15日
【特許出願人】
【識別番号】 000006792
【住所又は居所】 埼玉県和光市広沢2番1号
【氏名又は名称】 理化学研究所
【特許出願人】
【識別番号】 501293666
【住所又は居所】 東京都港区三田1丁目3番35号
【氏名又は名称】 株式会社ダナフォーム
申請人
【代理人】
【識別番号】 100088546
【住所又は居所】 東京都千代田区飯田橋4丁目5番12号 岩田ビル6階 谷川国際特許事務所
【氏名又は名称】 谷川 英次郎

次頁無

特願2002-235294

出願人履歴情報

識別番号 [000006792]

1. 変更年月日 1990年 8月28日
[変更理由] 新規登録
住 所 埼玉県和光市広沢2番1号
氏 名 理化学研究所

特願 2002-235294

出願人履歴情報

識別番号 [502049114]

1. 変更年月日 2002年 3月28日

[変更理由] 識別番号の二重登録による抹消

[統合先識別番号] 501293666

住所 東京都港区三田1丁目3番35号

氏名 株式会社ダナフォーム

特願 2002-235294

出願人履歴情報

識別番号 [501293666]

1. 変更年月日 2002年 3月28日
[変更理由] 識別番号の二重登録による統合
[統合元識別番号] 502049114
住所 東京都港区三田1丁目3番35号
氏名 株式会社ダナフォーム